VERİ MADENCİLİĞİ VE KARMAŞIK AĞ ANALİZİ TEKNİKLERİYLE SATINALMA SİPARİŞİ VERİLERİNİN ANALİZİ

Muhammet Şentürk
181402106

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Mehmet Ali Aksoy Tüysüz
İkinci Danışman: Dr. Öğr. Üyesi Volkan Tunalı

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Mart, 2020

VERİ MADENCİLİĞİ VE KARMAŞIK AĞ ANALİZİ TEKNİKLERİYLE SATINALMA SİPARİŞİ VERİLERİNİN ANALİZİ

Muhammet Şentürk
181402106
Orcid: 0000-0002-6393-4455

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Mehmet Ali Aksoy Tüysüz
İkinci Danışman: Dr. Öğr. Üyesi Volkan Tunalı

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Mart, 2020



JÜRİ VE ENSTİTÜ ONAYI

MUHAMMET ŞENTÜRK'ın "Veri Madenciliği Ve Karmaşık Ağ Analizi Teknikleriyle Satınalma Siparişi Verilerinin Analizi" başlıklı tezi 11.03.2020 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Maltepe Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği" nin ilgili maddeleri uyarınca Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans/~~Doktora/Sanatta Yeterlik~~ tezi oy birliğiyle/~~oy çokluğuyla~~, başarılı/~~başarısız~~ olarak kabul edilmiştir.

	Unvanı, Adı ve Soyadı	İmza
Üye (Tez Danışmanı)	Dr. Öğr. Üyesi Mehmet Ali Aksoy TÜYSÜZ	.....
Üye	Prof.Dr. Oruç Raif ÖNVURAL	.....
Üye	Dr. Öğr. Üyesi Selim BAYRAKLI	.....

Prof. Dr. Belma AKŞİT
Enstitü Müdürü V.

ETİK İLKE VE KURALLARA UYUM BEYANI

maltepe üniversitesi	LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ ETİK İLKE VE KURALLARA UYUM BEYANI	Doküman No	FR-178
		İlk Yayın Tarihi	01.03.2018
		Revizyon Tarihi	23.01.2020
		Revizyon No	01
		Sayfa	1

11/03/2020

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bulguların sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilmeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; çalışmanın Maltepe Üniversitesinde kullanılan "bilimsel intihal tespit programı" ile tarandığını ve öngörülen standartları karşıladığını beyan ederim.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

Muhammet ŞENTÜRK

TEŞEKKÜR

Yüksek lisans öğrenimim boyunca değerli vaktini ayırıp, bilgi ve tecrübeleri ile beni yönlendiren; her türlü kaynağını, ilgisini, desteğini ve yardımlarını benden esirgemeyen değerli hocalarım Dr. Öğr. Üyesi Mehmet Ali Aksoy Tüysüz'e ve Dr. Öğr. Üyesi Volkan Tunalı'ya;

Yüksek lisans eğitimim boyunca her zaman yanımda olan eşim İlknur Şentürk'e;

Bugünlere gelmem de çok büyük emekleri olan annem ve babama teşekkür ederim.

Muhammet Şentürk

Mart 2020

ÖZ

VERİ MADENCİLİĞİ VE KARMAŞIK AĞ ANALİZİ TEKNİKLERİYLE SATINALMA SİPARİŞİ VERİLERİNİN ANALİZİ

Muhammet Şentürk
Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Mehmet Ali Aksoy Tüysüz
İkinci Danışman: Dr. Öğr. Üyesi Volkan Tunalı
Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, 2020

Günümüzde şirketlerin büyüyen veritabanlarını, veri madenciliği ve karmaşık ağ analizi yöntemleri kullanarak analiz etmek kolay bir hal almıştır. Bu çalışmada kurumsal iki firmanın satınalma verileri iki farklı yöntem ile incelenmiştir. İlk yöntem; veri madenciliği başlığı altında anılan birliktelik kuralıdır. Burada, birlikte satın alınan ürünler birliktelik kuralı alt başlıkları olan sepet analizi yöntemi ve Apriori Algoritması uygulanarak incelenmiştir. Sepet analizi yönteminde yaygın olarak alınan ürünler ile en çok alınan ürün kümelerinin tespiti yapılmıştır. İkinci yöntem ise; karmaşık ağ analizi yöntemidir. Analiz ölçütlerinin kullanılması durumunda getireceği faydaların incelenmesi ve karşılaştırılmalı şekilde sunulması için faydalı bir analiz yapısına ulaşılması amaçlanmıştır. Bu çalışmada öncelikle elde edilen veriler içinden kirli veri olarak tarif edilen veriler ihmal edilmiş ve karmaşık ağ analizine uygun hale getirilmiştir. Sonrasında ise ağ modellenmesi yapılmıştır. Ağ analizi başlığı altında; görsel analiz, merkezilik analizi, topluluk analizi ve yapısal analiz yapılmıştır. Yapısal analizde de, satınalma ağının gerçek hayat ağlarına ait olup olmadığı rassal ağ ile karşılaştırılarak yapılmıştır. Son olarak ise sepet analizi raporlarında ön plana çıkmakta olan, madde kodlarının merkezilik analizi sonuçlarındaki değerleri incelenmiştir.

Anahtar Sözcükler: Birlikte Satınalınan Ürünler, Sepet Analizi, Apriori, KNIME, Karmaşık Ağ Analizi, Gephi

ABSTRACT

ANALYSIS OF PURCHASE ORDER DATA BY DATA MINING AND COMPLEX NETWORK ANALYSIS TECHNIQUES

Muhammet Şentürk
Master Thesis
Department of Computer Engineering
Computer Engineering Programme
Advisor: Asst. Prof. Dr. Mehmet Ali Aksoy Tüysüz
Co-Advisor: Asst. Prof. Dr. Üyesi Volkan Tunalı
Maltepe University Graduate School, 2020

Nowadays, it became easier to analyze companies' growing databases using data mining and complex network analysis methods. In this study, the purchasing data of two corporate firms were examined by two different methods. The first method is the association rule referred to under data mining. In this part, the products purchased together were examined by applying the basket analysis method and the Apriori Algorithm which they are subheadings of the association rule. In the basket analysis method, the most commonly purchased products and the most frequently purchased products were determined. The second method is a complex network analysis method. In the case of using analysis criteria, it is aimed to create a useful analysis format in order to examine the benefits and present them in a comparative way. In this study, first of all, the data described as dirty data were annihilated and made suitable for complex network analysis. Then, network modeling was performed. Under the title of network analysis, visual analysis, centrality analysis, community analysis and structural analysis were performed. In the structural analysis, it has been determined whether the purchase network belongs to real life networks or not by comparison with the random network. Finally, the values of the item codes that come to the forefront in basket analysis reports in the results of central analysis are examined.

Keywords: The Products Purchased Together, Basket Analysis, Apriori, KNIME, Complex Network Analysis, Gephi

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI ii
ETİK İLKE VE KURALLARA UYUM BEYANI iii
TEŞEKKÜR iv
ÖZ v
ABSTRACT vi
İÇİNDEKİLER vii
TABLOLAR LİSTESİ viii
ŞEKİLLER LİSTESİ ix
KISALTMALAR xi
ÖZGEÇMİŞ xii
BÖLÜM 1. GİRİŞ 1
BÖLÜM 2. İLGİLİ ÇALIŞMALAR 3
BÖLÜM 3. VERİ MADENCİLİĞİ 6
3.1.Veri Madenciliği 6
3.2.Tarihte Veri Madenciliği 7
3.3.Modeller ve Uygulama Alanları 8
3.3.1.Birliktelik Kuralları 9
3.3.2.Sepet Analizi 11
3.3.3.Apriori Algoritması 12
BÖLÜM 4. KARMAŞIK AĞ ANALİZİ TEKNİKLERİ VE SOSYAL AĞ ANALİZİ 15
4.1.Ağ 15
4.1.1.Ağ Kavramı 15
4.1.2.Ağ Kavramının Gelişimi 16
4.2.Sosyal Ağ Analizi 17
4.2.1.Neden Sosyal Ağ Analizi? 18
4.2.2.Sosyal Ağ Analizinin Gelişimi 19
4.2.3.Karmaşık Ağların Yapıları 20
4.2.4.Karmaşık Ağ Analizdeki Ölçütler 21
4.3.Ağ Seviyesinde Ölçütler 21
4.3.1. Topluluk Analizi 24
4.3.2. İki Parçalı (Bipartite) Graflar 24
BÖLÜM 5. I. YÖNTEM – BİRLİKTELİK ANALİZİ 26
5.1.Veriler ve Toplanması 26
5.2.KNIME Uygulaması ve Detayları 27
5.3.KNIME İndirilmesi ve Kurulumu 28
5.4.KNIME Uygulamasının Çalıştırılması 32
5.5.Çalışma Alanı Oluşturma 34
BÖLÜM 6. II. YÖNTEM – KARMAŞIK AĞ ANALİZİ 52
6.1.Karmaşık Ağ Analizi 52
6.2.Verilerin Karmaşık Ağ Analizine Uygun Hale Getirilmesi 53
6.3.GEPHI Uygulaması ve Detayları 57
6.3.1. Merkezilik Analizi 62
6.3.2. Topluluk Analizi 68
6.3.3. Yapısal Analiz 71
BÖLÜM 7. BULGULAR VE YORUMLAR 76
BÖLÜM 8. SONUÇ 78
KAYNAKÇA 80

TABLOLAR LİSTESİ

Tablo 5.1. Tez Çalışmasına Konu Olan Birinci Firmanın Veri Örneği 26

Tablo 5.2. Tez Çalışmasına Konu Olan İkinci Firmanın Veri Örneği 27

Tablo 5.3. GroupBy Node'u İle Tekilleştirilecek Sipariş No Örneği 39

Tablo 6.1. A ve B Firması Ham Verileri..... 53

Tablo 6.2. A firması verisi ve SQL sonuçları 55

Tablo 6.3. B firması verisi ve SQL sonuçları..... 56

Tablo 6.4. A Firmasının Derece Merkeziliğine Göre İlk 20 Madde Kodu 62

Tablo 6.5. B Firmasının Derece Merkeziliğine Göre İlk 20 Madde Kodu 63

Tablo 6.6. A Firmasının Arasındalık Merkeziliğine Göre İlk 20 Madde Kodu..... 63

Tablo 6.7. B Firmasının Arasındalık Merkeziliğine Göre İlk 20 Madde Kodu 64

Tablo 6.8. A Firmasının Yakınlık Merkeziliğine Göre İlk 20 Madde Kodu 64

Tablo 6.9. B Firmasının Yakınlık Merkeziliğine Göre İlk 20 Madde Kodu..... 65

Tablo 6.10. A Firmasının Özvektör Merkeziliğine Göre İlk 20 Madde Kodu 65

Tablo 6.11. B Firmasının Özvektör Merkeziliğine Göre İlk 20 Madde Kodu..... 66

Tablo 6.12. A firmasının birliktelik analizindeki madde kodlarının merkezilik analizi 67

Tablo 6.13. B firmasının birliktelik analizindeki madde kodlarının merkezilik analizi 68

Tablo 6.14. A Firması Karmaşık Ağında Bulunan Kümeler ve Özellikleri..... 69

Tablo 6.15. B Firması Karmaşık Ağında Bulunan Kümeler ve Özellikleri..... 69

Tablo 6.16. A ve B Firması Karmaşık Ağının Yapısal Ölçütleri..... 71

ŞEKİLLER LİSTESİ

Şekil 3.1. Veri Madenciliği Modelleri 9
Şekil 3.2. Apriori uygulamasına örnek 14
Şekil 4.1. İki düğüm ve yönsüz bir bağlantıdan oluşan ağ 15
Şekil 4.2. Dört Düğüm ve Yönlü Bağlantılardan Oluşan 16
Şekil 4.3. Königsber Köprüleri [34] 16
Şekil 4.4. Sosyogram Çalışması [33] 17
Şekil 4.5. Sociometrik star 19
Şekil 4.6. Sosyal ağın graf ve matris gösterimi [41] 20
Şekil 4.7. Sosyal Ağda Derece Merkeziliğine Örnek (degree centrality) [44] 22
Şekil 4.8. Sosyal ağda özvektör merkeziliğine örnek (eigenvector centrality) [44] 23
Şekil 4.9. Sosyal ağda arasındalık merkeziliği (betweenness centrality) [44] 23
Şekil 4.10. İki Parçalı Graf (Bipartite Graph) 25
Şekil 4.11. İki Parçalı Graf → Tek Parçalı Graf 25
Şekil 5.1. KNIME Ana Sayfa 29
Şekil 5.2. KNIME Eğitim Serileri 29
Şekil 5.3. KNIME Software Sayfası 30
Şekil 5.4. KNIME Uygulamasının Yapısı 31
Şekil 5.5. Programın ilk açılıştaki görüntüsü. 32
Şekil 5.6. KNIME Analytics Plarform – File Menü Ekranı 33
Şekil 5.7. Install KNIME Extensions ekranı 34
Şekil 5.8. Yeni çalışma alanı (Workflow) Açma 34
Şekil 5.9. Çalışma Alanı Oluşturma Sırasında Projeye İsim Verme 35
Şekil 5.10. KNIME Excel Reader (XLS) 35
Şekil 5.11. KNIME XLS Dosyasının Eklenmesi 36
Şekil 5.12. KNIME Excel Reader Node Configure 37
Şekil 5.13. KNIME Excel Reader Node Configure Refresh 37
Şekil 5.14. KNIME Excel Reader Configure ayarları sonrası çalıştırılması 38
Şekil 5.15. KNIME GroupBy Node Configure ayarları yapılmadan önce 39
Şekil 5.16. KNIME GroupBy Node Configure Grups ayarı 40
Şekil 5.17. KNIME GroupBy Node Configure Manual Aggregation ayarı 41
Şekil 5.18. KNIME GroupBy Node Configure ayarları sonrası 42
Şekil 5.19. KNIME GroupBy Node Group table seçim ekranı 42
Şekil 5.20. KNIME GroupBy Node Group Table 43

Şekil 5.21. KNIME Cell Splitter Node 43
Şekil 5.22. KNIME Cell Splitter Node Configure 44
Şekil 5.23. KNIME Cell Splitter Output View 45
Şekil 5.24. KNIME Item Set Finder Node ve Association Rule Learner Node 45
Şekil 5.25. Item Set Finder Node Configurasyon Ayarları 47
Şekil 5.26. A Firması Item Set Finder Node Sonuçları 47
Şekil 5.27. B Firması Item Set Finder Node Sonuçları 48
Şekil 5.28. Association Rule Learner Note Configure Ekranı 48
Şekil 5.29. A Firması Association Rule Learner Note Sonuçları 48
Şekil 5.30. A Firması Association Rule Learner Note Sonuçları - En Fazla Satınalınan Ürüne Göre Destek ve Güven 49
Şekil 5.31. A Firmasında Birlikte En Fazla Satın Alınan Ürünler 50
Şekil 5.32. B Firması Association Rule Learner Note Configure Ekranı 50
Şekil 5.33. B Firması Association Rule Learner Note Sonuçları 51
Şekil 5.34. B Firmasında Birlikte En Fazla Satın Alınan Ürünler 51
Şekil 6.1. Karmaşık Ağ Analizi Yöntemleri 52
Şekil 6.2. SQL üzerinde verilerin sosyal ağ analizine uygun hale getirilmesi 54
Şekil 6.3. Gephi adresinin giriş ekranı 57
Şekil 6.4. Gephi giriş ekranı 58
Şekil 6.5. A firması verilerinin Gephi programına eklenmesi 58
Şekil 6.6. B firması düğüm ve kenar sayısı 59
Şekil 6.7. A firması verisinin rasgele ağ yerleşimi 59
Şekil 6.8. B firması verisinin rasgele ağ yerleşimi 60
Şekil 6.9. A Firması Karmaşık Ağın ForceAtlas 2 Görünümü 61
Şekil 6.10. B Firması Karmaşık Ağın ForceAtlas 2 Görünümü 61
Şekil 6.11. A Firması Karmaşık Ağında Bulunan Kümeler 70
Şekil 6.12. B Firması Karmaşık Ağında Bulunan Kümeler 70
Şekil 6.13. A Firması Karmaşık Ağının Derece Dağılımı 72
Şekil 6.14. B Firması Karmaşık Ağının Derece Dağılımı 72
Şekil 6.15. A Firması Rassal Ağ Derece Dağılım Grafiği 74
Şekil 6.16. A Firması Rassal Ağ Analiz Sonuçları 74
Şekil 6.17. B Firması Rassal Ağ Derece Dağılımı Grafiği 74
Şekil 6.18. B Firması Rassal Ağ Analiz Sonuçları 75

KISALTMALAR

ABD : Amerika Birleşik Devletleri

CRM : Customer Relationship Management

KNIME : Konstanz Information Miner

ODBC : Open Database Connectivity

SAA : Sosyal Ağ Analizi

SDK : Software Development Kit

SQL : Structured Query Language

ÖZGEÇMİŞ

Muhammet Şentürk

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Y.Ls.	2020	Maltepe Üniversitesi, Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı
Ls.	2010	Anadolu Üniversitesi, İşletme Fakültesi, İşletme
Ön Ls.	2007	Anadolu Üniversitesi Bilecik Meslek Yüksek Okulu, Bilgisayar Teknolojisi ve Programlama
Lise	2005	Hereke Nuh Çimento Endüstri Meslek Lisesi, Bilgisayar Programlama ve Teknolojileri

İş/İstihdam

2011 – Halen Sistem Destek Kıdemli Uzmanı – Assan Bilişim A.Ş.

2007 - 2010 Teknik Servis Sorumlusu – Sarkom Bilgisayar Limited Şirketi

Kişisel Bilgiler

Doğum yeri ve yılı : Bulgaristan, 1987 Cinsiyet: E

Yabancı diller : İngilizce

e-posta :muhammedsenturk@hotmail.com

BÖLÜM 1. GİRİŞ

Geçmişten günümüze bir çok sektörde veriler dijital alanda saklanmakta ve saklanan verilerin sayısı her geçen gün giderek artmaktadır. Bu durum veri tabanlarının boyutunun büyümesine sebep olmaktadır. Bu sebeple veri tabanları üzerinde daha hızlı ve daha ucuz maliyetli işlemler yapabilmek adına çalışmalar hız kazanmıştır. Boyutu yüksek veri tabanlarında daha hızlı işlem yapılması ve maliyet azalması ile hem verinin kendisi ucuzlamış hem de veri saklama işlemi daha kolay hale getirilmiştir. Lakin saklanan her veri işe yararlılığı ve kullanılışlığı açısından değerlendirildiğinde;bir çok gereksiz verinin, veri depolarında saklanmakta olduğu görülmektedir. Veri depolarında saklanan verilerin, çok fazla olması sebebi ile bu veriler tek başına bir anlam ifade etmemekte ve verilerin çokluğundan dolayı istenilen bilginin alınması zorlaşmaktadır. Bu da istenmeyen bir durumu ortaya çıkarır. Ama bir çok veri analizi yöntemi sayesinde bu veri yığını, belirli bir amaç dogrultusunda sistematik şekilde işlenebilir. Değersiz görünen bir veri deposu, veri analizi yöntemleri sayesinde istenilen amaca yönelik çok degerli bilgilerin yer aldığı bir veri deposu halini alabilir [1].

Sepet analizi olarak bilinen perakende işlem verilerinin toplanması ve incelenmesi geçmişten günümüze giderek yaygınlaşmıştır. Örnek vermek gerekirse birçok süpermarket firması müşterileri için kartlar düzenlemiş ve düzenlemeye devam etmektedir [2, 3]. Bu kartlar müşterilere indirimler sunarken, süpermarket perakendecilerinin müşteri satın alma alışkanlıklarını daha iyi anlamasını sağlar. Böylece ürün yerleştirme kampanyaları, kişiselleştirilmiş pazarlama kampanyaları, ürün tanıtımlarının zamanlaması ve kapsamı gibi bir çok tasarım için önceden bilgi edinilmiş olmaktadır [3-5]. Kısaca pazar sepet analizinin görevi eyleme geçilebilir bilgi işlem veri tabanlarının keşfedilmesidir.

Ağ analizi ise; kendisine özgü analiz yaklaşımları, özel yazılımları, yöntemleri ve metrikleri bulunan ayrıca dünya genelinde çoğu verinin bir arada bulunduğu, birbirleriyle iletişim halinde olduğu ağ yapılarını incelemekte olan bir disiplindir.

Bu tez çalışmasında, kurumsal iki firmanın satınalma verileri birliktelik kurallarından sepet analizi yöntemi ile incelenmiş ve karmaşık ağ analizi yöntemleri ile analiz edilmiştir. Sepet analizi yönteminde apriori algoritması kullanılmıştır. Ağ analizi

yönteminde ise veriler öncelikle iki parçalı graflar yöntemi ile ağ analizine uygun hale getirilmiştir. Verilerin ağ analizine uygun hale gelmesi sonrası, Gephi programı kullanılarak; merkezilik analizi, topluluk analizi ve yapısal analiz uygulanmıştır. Sepet analizi ve ağ analizi yöntemlerinde çıkan sonuçlar karşılaştırılarak değerlendirme yapılmıştır.

BÖLÜM 2. İLGİLİ ÇALIŞMALAR

Tunalı ve Güvenoğlu'nun 2019 yılında yaptığı bir çalışmada, karmaşık ağ analizi yöntemleri kullanılarak öğrencilerin yükseköğretim bölümü tercihleri analiz edilmiştir. Öğrenci tercih verileri, geliştirmiş oldukları web sayfası tarama motoru kullanılarak, Yükseköğretim kurumu tarafından sağlanmış olan Yök Atlas veri portalı içerisinden toplanmıştır. Toplanan ana veriden, birlikte tercih edilme ağı olarak adlandırılan 622 düğüm ve kenara sahip birliktelik ağı oluşturulmuştur. NodeXL ile Cytoscape programları kullanılarak bu ağda keşif olarak detaylı karmaşık ağ analizi gerçekleştirilmiştir. Farklı düğüm merkezilik analiz ölçütleri kullanılarak öğrencilerin diğer bölümlerle beraber çoğunlukla seçmiş olduğu en popüler bölümler tespit edilmiştir. Ayrıca çeşitli topluluk tespiti analiz yöntemleri kullanılarak ağda yerleşik bölüm kümeleri gözlemlenmiştir. Son olarak ise; ağa karşılık gelmiş olan rastgele ağ ile karşılaştırılarak bir yapı analizi gerçekleştirilmiş ve ağın çoğunun gerçek hayat ağının verdiği ortak karakteristik özelliğine sahip olduğu gösterilmiştir [6].

Reader ve Chawla'nın 2010 yılında yaptığı çalışmada, Birliktelik kurallarından sepet analizi yöntemi kullanarak satınalma verileri üzerinde analizler gerçekleştirilmiştir. Bu çalışmada veriler bir ürün ağı olarak modellenerek, verilerden anlamlı topluluklar ve kümeler meydana getirilmiştir. Ağ tabanlı yaklaşım ile ürünler arasındaki bağın daha düzgün bir şekilde analiz edilmesi gerçekleştirilmiştir. Bu sayede birlikteliklerin daha anlamlı verilerde aranmasıyla, istenilen veriye daha kısa sürede ulaşılması sağlanmıştır. Son olarak hem geleneksel hem de ağ tabanlı teknikler birleştirilerek kapsamlı bir analizi yapılmıştır [3].

Das'ın 2012 yılında yaptığı bir çalışmada, ağ tabanlı ürün ağ analizi önerilmektedir. Ürünler arası ilişkiyi inceleyen sepet analizi ile karşılaştırıldığında ise ürün ağ analizi, ürünler arası ilişkilerin ağ seviyesinde genişletilmiş bir bakış açışı vermektedir. Bu amaçla, sepet analizi ve birlikte satın alınan ürün ağları önerilmektedir. Topolojik özelliklerini ve bu ağların yapılarını analiz etmek için iki ağ karşılaştırma yapılarak değerlendirilir. Yapılan değerlendirmede sepet analizinin daha etkin ve verimli çalıştığı tespiti yapılmıştır [7].

Demirok'un 2018 yılında KNIME programını kullanarak yapmış olduğu çalışmada; internet üzerinden satış yapan firma, veri madenciliği yöntemleri ile birlikte alınan ve en çok tercih edilen ürünlerin analizini yapmış ve sonuçlarını incelenmiştir. Bu çalışmada birliktelik kuralı algoritması olan Apriori algoritması kullanılmıştır. Birliktelik kuralı, birlikte alınan ürünlerin hangi sıklıkla alındığını belirleyen ve analiz eden bir yöntemdir. Müşterilerin kullanım alışkanlıklarına ve karakteristik özelliklerine göre hangi müşterinin, hangi ürünleri, ne zaman aldığı analiz edilerek çıkan sonuçlar değerlendirilmiştir. Yapılan bu çalışmada iki farklı amaç bulunmaktadır. Bunlar müşterilerin en çok tercih ettiği ürün grupları ve en çok tercih edilen ürünlerin tespitidir [8].

Çelik KNIME programını kullanarak 2015 yılında yaptığı çalışmada, veri madenciliği yöntemlerini kullanarak yolcular için en ilgi çekici reklam ve anketin hangileri olduğu ile ilgili analiz yapmıştır. Bu çalışma içerisinde, Karar Ağaçları, K en yakın komşu ve sınıflandırma yöntemleri uygulanmıştır. Veriler sentetik yol ile çoğaltılıp elde edilmiş ve veri seti üzerinde uygulanarak 3 farklı kategoride sınıflandırılmıştır. Karar Ağaçları, K en yakın komşu ve sınıflandırma yöntemleri ile seçilmiş olan anket ve reklam değişikliğinde başarılı sınıflandırma bulunmuştur. Çıkan sonuçlar, yöntemler bazında kıyaslanarak araç-eğlence sistemlerinde en ideal yöntem sağlanmıştır [9].

Cavieres ve Rios'un 2014 yılında yaptığı bir çalışmada; süpermarket gibi büyük marketler ve şirketlerdeki en büyük sorunun satılan ürünlere ait birliktelikleri bulmak olduğu belirlenmiş ve mevcut tek bilginin, satış tarihleri olduğu tespit edilmiştir. Bu tespitlerden yola çıkarak, sepet analizi teknikleriyle milyonlarca karmaşık veri setini analiz edebilen grafik madenciliği yöntemlerine dayanan yeni bir yöntem geliştirilmiştir. Klasik bir yaklaşım ile karşılaştırma yapılarak sırasıyla 238.000.000 ve 128.000.000 işlem gerçekleştiren bir süpermarket zincirinde yapılan çalışmanın etkinliği gösterilmiştir [10].

Bolata Gephi programı ile 2018 yılında yapmış olduğu çalışmada, ülkeler arası göç ağı oluşturmuş ve karmaşık ağ analizi yöntemleri ile bu ağı analiz etmiştir. Veriler, Birleşmiş Milletler'e ait "Trends in International Migrant Stock: Migrants by Destination and Origin" isimli çalışmasından elde edilerek uluslararası göç verileri kullanılmıştır. Bu veriler öncelikle gürültü ve kirli veriden arındırılmıştır. Ağın modellenmesi sonrası

ülkeler arası göç ağı çıkarılmıştır. Analiz türleri olarak, ağ üzerinde görselleştirme ile görsel analiz, merkezilik analizi ve topluluk analizi meydana getirilmiştir. İlgili analizlerden sonra ise yapısal analizde, ülkeler arası göç ağının gerçek dünya ağlarına ait karakteristik özelliklere sahip olup olmadığı sonucu aranmıştır [11].

BÖLÜM 3. VERİ MADENCİLİĞİ

3.1.Veri Madenciliği

Veri madenciliği; mevcut veriler içerisinde görünmeyen, anlaşılmayan, öncesinde bilinmeyen ama potansiyel olarak kullanışlı bilgilere ulaşılmasıdır. Diğer bir değişle; verileri özetleme, kümeleme, yapılan değişikliklerin analizi, sapmaların tespiti gibi çeşitli teknik yaklaşımlar içerir [12].

Veri madenciliği biraz daha indirgendiğinde bilgisayar biliminin disiplinler arası ilişkisi olarak belirtilebilir. Veri tabanlarında bilginin bulunması (Knowledge-Discovery in Databases) olarak da karşımıza çıkabilir [13].

Veri madenciliği; yapay zeka, makine öğrenmesi, istatistik, veri tabanı sistemleri yöntemlerinde ve büyük miktardaki veri setlerinde örüntülerin keşfinde kullanılan bir süreçtir. Veri madenciliğinin asıl amacı bir veri setinden bilgiyi üretmek ve onu daha ileri kullanımlar için anlaşılır net bir veriye dönüştürmektir [13]. Son yıllarda veri madenciliğinin kullanımı oldukça yaygınlaşmıştır.

Veri madenciliğinin çalışma prensibi, veriler arasında bağlantıları bulmak için büyük miktardaki veri tabanı üzerinde çalışan hesaplama gücünü kullanır. Siber güvenlik ve ulusal güvenliği sağlama uygulamalarında da kullanılır. Bunlar dışında veri madenciliği; şüpheli insanların bulunmasında, yapılan iş ve işlemlerin daha verimli olarak ele alınabilmesinde, yeni müşterilerin kazanılmasında, müşterilerin satın alma davranışlarının incelenmesinde kolaylık sağlamaktadır. Ayrıca biyometrik uygulama ve kredi kartı dolandırıcılığında da yoğun olarak kullanılmaktadır [14]. Yukarıda bahsedilen konular dışında; veri depolama birimleri, barkod okuyucular, manyetik kartlar ve RFID teknolojisine paralel olarak gelişim göstermekte olan veri madenciliği kullanımı gittikçe yaygınlaşmakta olan bir alan haline gelmiştir [15]. Önceden bilinmeyen veya öngörülmeyen sonuçlara en dikkat çeken örneği ise veri madenciliğinde ‘bilinmeyenin’ şaşırtıcı bir şekilde ön plana çıkarak çocuk bezi ile biranın beraber alınma örneğidir. ABD’de bulunan büyük bir mağaza zincirinin yapmış olduğu araştırma sonucuna göre çocuk bezi ve bira satışlarının özellikle de cuma akşamları olmak üzere çok fazla beraber alındığı tespit edilmiştir. Çocuk bezini satın alan kişilerin de büyük bir bölümü bununla birlikte bira da satın aldıkları görülmüştür. Kısacası, yeni baba olmuş kişilerin hafta

sonları dışarıya çıkmamak adına kendileri için de alışveriş yaptıkları sonucuna ulaşılabilmektedir [16].

3.2.Tarihte Veri Madenciliği

1950 yılları içerisinde geliştirilmiş olan ilk bilgisayar basit işlemler ile basit aritmetik işlemleri yapabilmek için tasarlanmıştır. 1960'larda ise ilk veritabanı ile verilerin korunması kavramı gelmiştir. 1960 sonlarındaysa bilim adamları veritabanlarını depolama amacı dışında işlem yapabilme kapasitesinden dolayı da kullanmaya başlanmıştır. Ayrıca bilim adamları kolay öğrenilebilecek algoritmaların geliştirmesiyle bu yapıları kullanılabilir seviyeye ulaştırmışlardır [17]. 1970 yıllarındaysa E.F. Codd'un yayınlamış olduğu ilişkisel veritabanları ile ilgili bir makalede, insanlar ilişkisel veritabanları ile tanışmış oldular. Bununla birlikte, bilim adamları basit anlamda makine öğrenimi gerçekleştirmeye başlamıştır. 1980 yıllarında SQL yazılımı çıkarılması ile veritabanı yönetim sistemlerinin kullanımı artmıştır. Kamu, özel sektör ve bilimsel alanlar başta olmak üzere kullanılmaya başlanmıştır. Firmalar bu yıllarda ürünlerin, müşterilerin ve tedarikçilerin bilgilerini veritabanlarında tutmaya başlamışlardır. Bu verilerin mevcut olduğu veritabanları sayesinde büyük verileri saklamaları mümkün olmuştur. Bu yıllarda ilk olarak A-CRM projelerinin temelini oluşturan veritabanları yapılmıştır [18]. 1990 yıllarında yazılım dünyasında Visual Basic, ODBC, Excel ve Access gibi programlar yer almaya başlamıştır. Yine aynı yıllar içerisinde büyümekte olan veritabanlarından istenilen bilgilere sadece sorgu ve raporlar ile ulaşılabileceği anlaşılmıştır. Değişik yöntemler bulabilmek adına uluslararası çalışmalar yapılmıştır [17].

2000'li yıllarda ise veri madenciliği tüm alanlarda kullanılmaya başlanmıştır. Büyük şirketler; Twitter, Facebook, Google ve Linkedin ile kamu sektöründeki bazı firmaların bu alanda yatırım yapmasıyla diğer şirket ve kuruluşlarda ilgi ve güven doğurmuştur [17]. Artık veri madenciliği günümüzde birçok alanda uygulanmaktadır.

Kısaca özetlemek gerekirse; veri madenciliği, her gün gelişmekte olan teknolojinin sonucu olarak günümüzdeki yerini almıştır. İnsanlık tarihi boyunca veriler toplanmış ve toplanmaya da devam etmektedir. Toplanmış olan bu veriler yorumlanmış

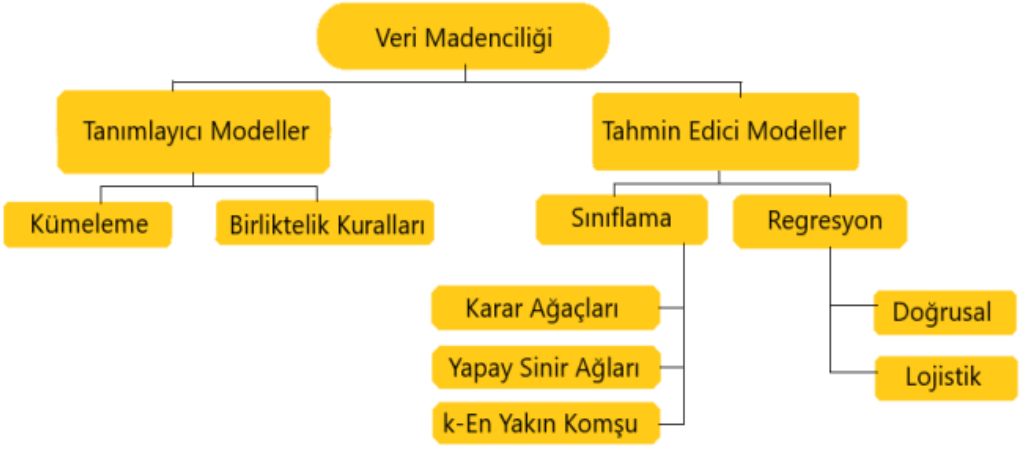
ve verilerden önemli, yararlı bilgiler elde edilmiştir. İşlemlerin yapılabilmesi için çeşitli donanımlar ve yazılımlar geliştirilmiştir. Bu sayede veriler ve bilgiler günümüze kadar gelmiştir.

3.3.Modeller ve Uygulama Alanları

Veri madenciliği modelleri iki ana başlık altında incelenmektedir. Bunlar tanımlayıcı (Descriptive) ve tahmin edici (Predictive) modellerdir [1, 19]. Tahmin edici (Predictive) modellerde, sonuçları belli olan verilerden yola çıkılarak yeni model geliştirilmektedir. Geliştirilmiş olan bu modelden faydalanılarak sonuçları bilinmeyen verilerin sonuç değerlerinin tahmin edilmesi hedeflenmiştir. Örnek olarak bir banka eski yıllarda verdiği kredilere ait tüm bilgilere sahiptir. Bu veriler içerisinde bağımsız değişkenler kredi almış olan müşterilerin özelliklerini, bağımlı değişken değeriyse kredilerin geri ödenip ödenmediğini belirtir. Toplanan bu veriler sayesinde müşterilerin bir daha ki kredi talebinde verilecek olan kredi tutarının geri ödenip ödenmeyeceği tahmini ele alınmaktadır. Tanımlayıcı modellerdeyse karar vermede ve rehberlik yapmada değerlendirilebilecek mevcuttaki veriler içerisindeki örüntülerin tanımlanması yapılmaktadır [1]. Aylık geliri X-Y aralığında olan ve arabası bulunan çocuklu aileler ile çocuğu bulunmayan ve geliri X-Y aralığından az olan ailelerin alım güçleri birbirlerine benzerlik gösterdiği ifade edilmektedir. Bu gibi örnekler tamamlayıcı modellerdir [20].

Tahmin edici modellerde şu sorunun yanıtını ararlar: “Bu işlemde dolandırıcılık var mıdır?" ya da “Bu müşteriden ne kadar gelir elde edilebilir?"

Tanımlayıcı bir model, veriler arasındaki gizli kalmış ilişkileri meydana çıkarır ve tipik olarak şöyle bir sonuç elde edilir: “Çocuk bezi alan bir müşterinin mama alma olasılığı diğerlerinden 3 kat daha fazladır.” [17].



Şekil 1 Veri Madenciliği Modelleri

3.3.1.Birliktelik Kuralları

Veri madenciliğinin ilk tekniklerinden biri olan birliktelik kuralları analizi 1993 yılında Agrawal, Imielinski ve Swami tarafından ilk defa kullanılmıştır [21]. Birliktelik Kuralları; satın alımların beraber olma olasılığını inceleyen veri madenciliği yöntemlerine denir. Bu metotlar sayesinde birlikte olma kurallarını belirli olasılıklarla vermektedir. Diğer bir değişle geçmişteki verilerin analiz edilmesiyle, çıkan sonuçlardan birliktelik tespiti yapılarak ileriye dönük planlamalar yapılmasını sağlayan bir tür yöntemdir.

Birliktelik kuralı ile yapılmak istenen; alışverişler sırasında kişilerin almış oldukları ürünler içerisindeki birliktelikleri çıkarmak, bu birliktelik verileriyle kişilerin almış oldukları ürünlerdeki alışkanlıklarını bulmaktır. Satıcı firmalar, keşfedilmiş olan bu birliktelikler ile verimli ve etkili pazarlama stratejisine sahip olmaktadır. Analizi daha da genişletirsek; bir marketten, yumurta ve süt almış olan kişilerin %70'i bunlarla beraber yoğurt da almıştır. Bu tarz birlikteliklerin çıkarılabilmesi için örüntü içinde bulunan ürünlerin birden fazla beraber alınmış olması gerekmektedir. Veri çok fazla olduğunda, birliktelik kuralları için kullanılan algoritmaların hızlı olması gerekmektedir [22]. Bu birliktelik kuralı uygulamasına ait en etkili ve hızlı algoritmalardan biri olarak sepet analizine örnek verilebilir [23].

Birliktelik kuralında kullanılan başlıca algoritmalar: Apriori, FP-Growth, GRI, Eclat, Carma, Sequence, ve diğerleri. Burada en çok kullanılmakta olan ve tercih edilen algoritma Apriori algoritmasıdır.

Birliktelik kuralını matematiksel formül şeklinde ifade etmek istersek; $A \Rightarrow B$ kuralındaki destek değeri (%s) veriler kümesinde $P(A \cup B)$ olasılığı yani A ve B elemanının bir veri kümesinde birlikte olma olasılığını, güven değeri (%c) ise P(B|A) olasılığı yani veri kümesinde A'nın olması halinde B'nin olma olasılığını ifade etmektedir.

Agrawal, Imielinski ve Swami tarafından bulunan birliktelik analizinin matematiksel modeli, $I = \{i_1,\ i_2, \dots, i_m\}$ kümesine "ürünler" olarak ifade edilmektedir.

D, verinin tamamındaki hareketleri simgeler, T ise ürünlerin tüm hareketini belirtir. TID ise her harekette bulunan tek belirteçtir.

Farklı bir değişle Birliktelik kuralı;

$$A_1, A_2, \dots\dots, A_m \ \Rightarrow B_1, B_2, \dots\dots, B_n$$

olarak tanımlanabilir. $i = 1,2, \dots, m$ ve $j = 1,2, \dots, n$ olarak yazdığımızda özetlemek gerekirse;

$$A_i = B_j$$

Yukarıda bulunan A_i ve B_j yapılan nesneler veya işlerdir. Bu tanım içerisinde "$A_1, A_2, \dots\dots, A_m$" nesneler veya işler olduğunda yoğun olarak "$B_1, B_2, \dots\dots, B_n$" nesnelerin veya işlerin aynı hareket veya olay içerisinde bulunduğunu gösterir [24].

Birliktelik kuralının minimum değeri (Mindes) kişi tanımlanmış güven ve destek eşik değerlerini sağlayacak şekilde geliştirilir Ürün kümelerindeki destek değeri D ile gösterilmektedir ve tüm hareketler içerisinde ilgili ürün kümelerini belirten hareketlerin yüzdesini ifade eder. A ile B ürün kümelerinin birliktelik kuralı "$A \Rightarrow B$" olarak belirtilirse destek değeri aşağıda bulunan şekilde olur;

$$Destek\ (A \Rightarrow B) = \frac{A\ ve\ B'nin\ bir\ arada\ yer\ aldığı\ kayıtların\ sayısı}{Toplam\ kayıt\ sayısı} \qquad (4.1)$$

$A \Rightarrow B$ olduğunda birliktelik kuralı güven değeri, A'yı belirten hareketlerin B'yi de belirtme yüzdesidir. Örneğin, kural % 85 güvenirliğe sahipse, A'yı belirten ürün kümesinin % 85 B'yi de ifade etmektedir. İşlere bağlı bulunan veri satırları belirtilmişse, $(A \Rightarrow B)$ güveni aşağıdaki gibi ifade edilir.

$$Güven\ (A \Rightarrow B) = \frac{A\ ve\ B'nin\ bir\ arada\ yer\ aldığı\ kayıtların\ sayısı}{A'nın\ yer\ aldığı\ kayıtların\ sayısı} \qquad (4.2)$$

Güven değeri % 100 olduğunda, kural tüm analizlerde doğrudur ve bu gibi kurallara "kesin" denmektedir. Birliktelik kuralları için geliştirilmiş olan başlıca algoritmalar bunlardır; AIS, Apriori, RARM - Rapid Association Rule Mining, CHARM, SETM [21, 22, 25-28]. Bu algoritmalardan, ilk olanı AIS iken, çok bilinense Apriori algoritmasıdır.

3.3.2.Sepet Analizi

Sepet analizi, herhangi bir mağazanın satınalma verilerinden birlikte alınmış olan ürünleri ele alarak satınalma modellerini bulmaya odaklanmış veri madenciliği modellerinden biridir. Diğer bir değişle; müşteriler farklı zamanlarda, farklı miktarlarda, farklı ürün veya ürün grubu satın almıştır. Sepet analizi bir müşterinin satınalmış olduğu ürünleri hangi ürünler ile beraber tercih ettiğine dair verileri incelememizi sağlar [7]. Sepet analizi, işlem verileri üzerinde en çok uygulanan tekniklerden biridir. Veri madenciliği tekniklerinin geniş ailesinin bir parçasıdır. Sepet analizinin amacı, bir müşterinin iki farklı ilkeye dayanarak daha fazla para harcamasını sağlamaktır. Birincisi, aynı ürünün büyük bir miktarını satın almak veya yeni özellikler eklemek olan Up-Selling'dır. İkinci yol, farklı kategorilerden daha fazla ürün eklemekten oluşan çapraz satış tekniğidir [4, 21].

Sepet analizinin özeti ve kullanım amacını indirgememiz gerekirse; müşterilerin aynı anda hangi ürünleri satın aldıklarına karar vermek için alışveriş ürünlerinin, alışveriş

sepetlerine veya alışveriş listesine yerleştirilmesinde müşterilerin davranışlarını gözlemlemek ve bu veriler ile sonuçlar üretmektir. Bir müşterinin alışveriş sepet modelini tanımlamak, bir şirketin iş stratejisi ve gereksinimlerinin belirlenmesinde yardımcı olabilecek bir unsurdur [29].

3.3.3.Apriori Algoritması

Apriori Algoritması, verileri bir önceki adımdan alındığından "prior" anlamına gelen Apriori olarak adlandırılmıştır [4]. Algoritmanın temelinde tekrarlayan (iteratif) bir ifadesi bulunur ve hareket bilgileri içermekte olan veri tabanlarında sık olan öğe kümelerinin bulunmasında kullanmaktadır. Agrawal tarafından 1994 yılında geliştirilmiş olan Apriori algoritması, veri madenciliği tarihinde birliktelik kurallarının bulunması adına geliştirilmiş büyük bir buluştur [4].

Apriori algoritması veri setindeki $(k+1)$ satırı araştırmak için düz bir arama olarak bilinen yinelemeli bir yaklaşımı kullanır. İlk olarak 1-öğe kümesi, her bir 50 öğenin minimum destek kriterini sağlayan öğelerini toplamak için veri tabanını tarar. Ortaya çıkan set L_1 olarak adlandırılır. Daha sonra 2-öğe kümesi bulmak için L_1 seti kullanır ve set L_2 olarak adlandırılır. L_3'ü bulmak içinse L_2'yi kullanır. Her L_k'nın bulunması için veri tabanı taranır. Süreç sık geçen öğe kümesi bulunmadığında sonlanır [30].

Apriori algoritması genellikle alışveriş uygulamalarında kullanılması sebebiyle sıklıkla sepet analizi olarak bilinmektedir. Bir veri kümesi içerisindeki nesnelerin birbirleri ile ilişkisi destek ve güven değerleri ile hesaplanır. Destek bir veri kümesi içinde ilişkinin ne kadar yoğun olduğunu, güven değeriyse A nesnesini almış bir müşterinin hangi olasılıkla B nesnesini de alacağını ifade etmektedir. Destek ve güven kriterlerinin her biri ne kadar büyükse ilişkide o kadar büyük demektir [31]. Destek ve güven kriterleri aşağıdaki denklemler ile hesaplanır.

$$Destek\ (A) = \frac{A \text{ öğe } kümesi\ sayısı}{Toplam \text{ öğe } küme\ sayısı} \tag{4.3}$$

$$Destek\ (A,B) = \frac{(A,B) \text{ öğe } kümesi\ sayısı}{Toplam \text{ öğe } küme\ sayısı} \tag{4.4}$$

$$Güven\ (A.B) = \frac{(A,B)\ \text{öğe kümesi sayısı}}{A\ \text{öğe küme sayısı}} \tag{4.5}$$

$$Güven\ (A,B) = \frac{Destek\ (A,B)}{Destek\ (A)} \tag{4.6}$$

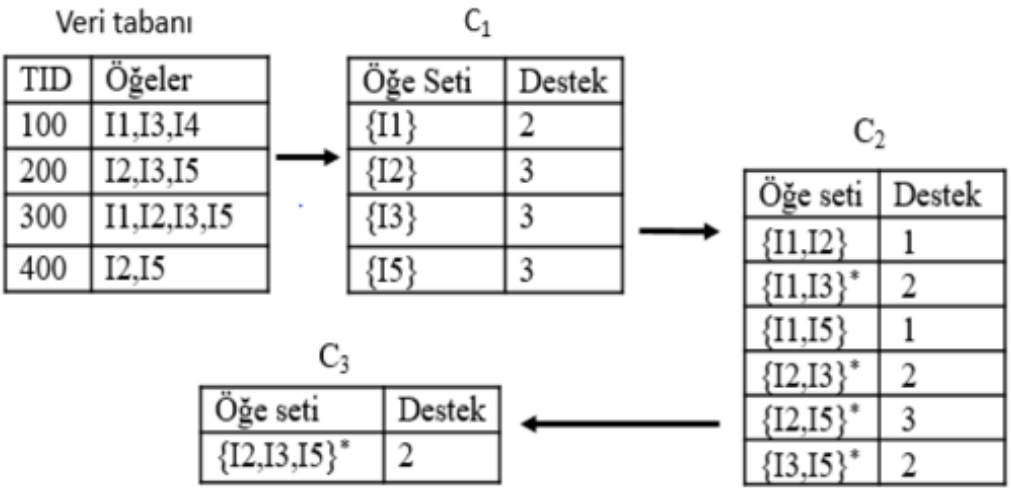
Apriori algoritmasının en bilinen özelliği bir öge kümesi yaygın öge kümesi ise bu öge kümesinin tüm alt kümeleri de yaygın öge kümesi olmalıdır [32].

Algoritma k. iterasyonda ($k \geq 2$ için) $(k-1)$ yaygın öge kümesine dayanan k yaygın öge kümesi adayı oluşturmakta ve k adet yaygın öge setinden oluşan tam seti (L_k) bulmak için veritabanını bir kez daha taramaktadır. (L_k)'yi bulmak için birleştirme adımı ve budama adımı olarak adlandırılan iki adımlı bir işlem uygulanmaktadır. Birleştirme adımında, (L_k)'yi bulmak için k adet aday öge kümesinden oluşan bir set $L_k - 1$ ile birleştirilir. Bu aday set C_k ile gösterilir. C_k, L_k'nın bir üst kümesi olup üyeleri yaygın olabilir veya olmayabilir ancak tüm yaygın k öge kümesi C_k'da bulunur. C_k'daki her bir adayın uyumunu belirlemek için yapılan veritabanı taraması L_k'nın belirlenmesiyle sonuçlanır. C_k'nın büyüklüğü arttıkça hesaplama zamanı da artış gösterir. C_k'nın büyüklüğünü düşürmek için budama adımı uygulanır. Bu adımda:

- Yaygın olmayan herhangi bir $(k-1)$ öge kümesi bir yaygın k-öge setinin alt kümesi olamaz.

- Eğer bir aday k öge kümesinin herhangi bir $(k-1)$ alt kümesi $L_k - 1$'de yoksa aday yaygın öge kümesi olamaz ve C_k 'dan kaldırılır.

Bu yüzden bu algoritma destek ve güven eşik değeri koşullarını sağlayan bir veritabanındaki büyük öge kümelerini keşfetmeye uygun bir algoritmadır. Yaygın küme olma durumunda olmayan birçok kümeyi budayarak yaygın öge kümelerini bulmak için kullanılan bir yinelemeli yöntemdir. Algoritmanın kısıtlamaları arasında çok sayıda aday öge kümeleri üretilebilmesi ve n sayıda (boş olmayan öge kümeleri sayısı) veritabanı taraması gerektirmesi yer almaktadır. Ayrıca keşfedilen kuralların sayısı çok fazla olup bunların büyük bölümü de ilginç olmayabilir. Şekil 3.2.'de Apriori algoritmasına örnek bulunmaktadır.



Şekil 2 Apriori uygulamasına örnek

BÖLÜM 4. KARMAŞIK AĞ ANALİZİ TEKNİKLERİ VE SOSYAL AĞ ANALİZİ

4.1.Ağ

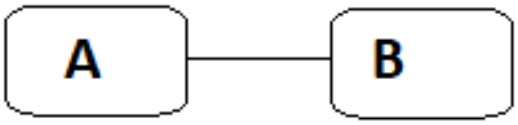
Sosyal hayatta; bilgisiyar ağı, ulaşım ağı, protein ağı, iletişim veya trafik ağı gibi birçok yapay ve doğal ağlardan oluşan bir dünyada yaşanmaktadır. Ağlar, günümüzde yeni bir değerler dizisi olarak ifade edilmektedir. Ağların türleri, dirençleri, bilgiyi nasıl yaydıkları veya engelledikleri son zamanlarda tartışılmaya başlanan sorulardır.

4.1.1.Ağ Kavramı

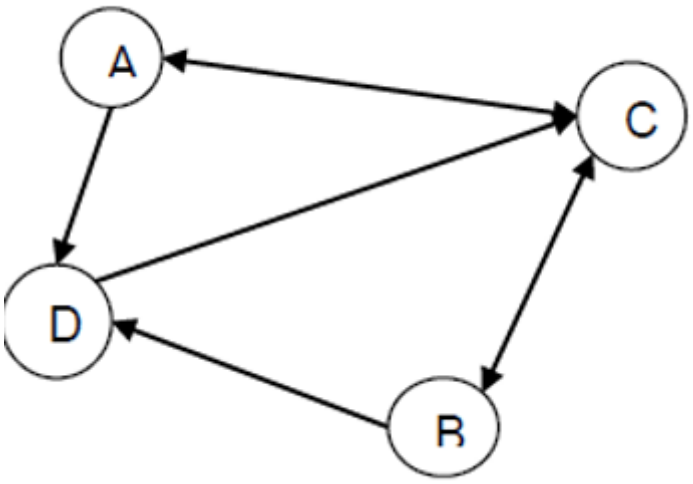
Ağlar günümüzde sistemi oluşturan nesnelerin, karşılıklı etkileşimiyle beliren karmaşık sistemleri göstermenin farklı bir bakış açısıdır [33].

Ağlar, düğümler ile bu düğümlerin arasındaki bağlantılardan meydana gelmektedir. Bir ağdan bahsedebilmek için bağlantılı en az iki düğüm gerekmektedir. Ağ da bulunan düğümler için çeşitli ağ analizi programlarında birim, köşe (vertex); bağlantılar için ise kenar (edge) gibi terimler kullanılmaktadır. Herhangi bir ağdaki düğümlerin çok olması ağın büyük olduğunun göstergesidir [33].

Düğümler arasındaki bağlantı yönüne ilişkin bir ok olmadan gösterilmiş ise yönsüz bağlantı (undirected edge) ya da simetrik bağlantı; düğümlerin arasındaki bağlar yönlü oklarla gösterilmiş ise yönlü bağlantı (directed edge) ya da asimetrik bağlantı olarak adlandırılmaktadır [33].



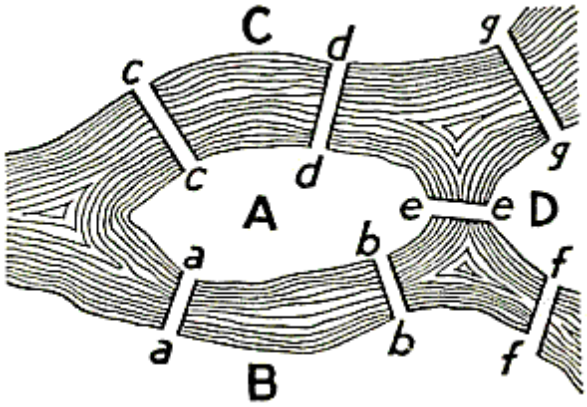
Şekil 3 İki düğüm ve yönsüz bir bağlantıdan oluşan ağ



Şekil 4 Dört Düğüm ve Yönlü Bağlantılardan Oluşan

4.1.2.Ağ Kavramının Gelişimi

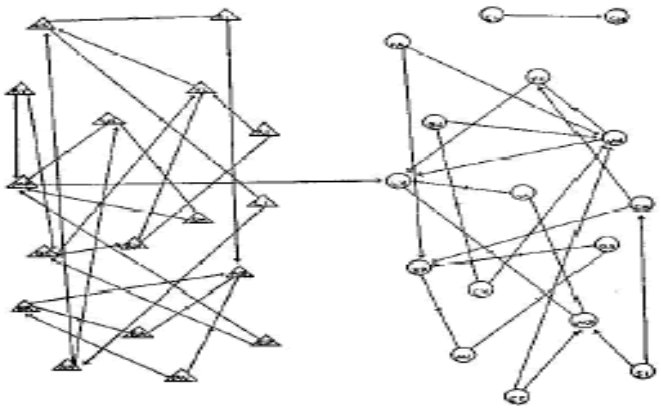
Ağ biliminin temeli; 1736 yılında matematikçi Leonhard Euler'in, Prusya kenti Königsberg'de bulunan Pregel nehri üzerindeki köprüler ile ilgili makalesine dayanmaktadır. Bu makalenin açıklamaya çalıştığı soru: "herhangi bir köprüden birden fazla geçmeden, yedi köprünün tamamını geçmek için bir patika oluşturulabilir mi?" sorusudur. Şekil 4.3.'de Königsber köprüleri bulunmaktadır [34].



Şekil 5 Königsber Köprüleri [34]

Leonhard Euler, Königsberg'deki kara parçalarını düğümler, nehir üzerindeki yedi köprüyü de bu düğümler arasındaki bağlar olarak tanımlamıştır. Çizimin ardından bir ispat yapmış ve soruya "Hayır" cevabını vermiştir. Matematikte Graf Kuralının doğuşu olarak adlandırılan bu soru ; çizim, düğümler ve bu düğümler arasındaki bağlardan oluşan ilk ağ çizimi olarak görülmektedir [33].

Euler'den sonra ağ bilimi ile ilgili çalışma 1933 yılında Jacob Moreno tarafından yapılan, ilkokul öğrencilerinin sosyal yapılarını açıklayan sosyogram çalışmasıdır. Bu çalışmada Moreno, bir okuldaki kız ve erkek öğrencilerin arasındaki bağları incelemiştir. Bu okuldaki erkek öğrenciler sadece erkek öğrencilerle, kız öğrenciler ise sadece kız öğrencilerle arkadaşlık etmektedir; ancak bir erkek öğrenci bir kız öğrenciye karşılıksız ilgi duymaktadır. Bu bağların incelenmesi sonucunda Şekil 4.4.'deki ağ ortaya çıkmıştır. Moreno'nun sosyogram çalışması sosyal ağ analizi çalışmalarının gelişmesinde önemli bir rol oynamıştır [33].



Şekil 6 Sosyogram Çalışması [33]

4.2.Sosyal Ağ Analizi

Sosyal ağ analizi (SAA), 1960 ve 1970'li yıllarda, ağırlıklı olarak sosyal psikoloji içinde, sosyolog ve araştırmacılar tarafından geliştirilmiştir. Matematik, istatistik ve bilgisayar ile işbirliği içinde daha da geliştirilmesiyle ekonomi, pazarlama ve endüstri mühendisliği gibi diğer disiplinlerde, cazip bir araç olarak analiz tekniklerinin hızlı gelişmesine yol açmıştır [35]. SAA çağdaş sosyolojinin bir tekniği olarak ortaya

çıkmıştır. SAA, düğümler ve bağlardan (kenarlar, bağlantılar veya ilişkiler olarak da adlandırılır) oluşan ağ teorisi terimi içinde sosyal ilişkileri incelemektedir. Düğümler ağ içindeki bireysel aktörlerdir ve bağlar aktörler arasındaki ilişkilerdir. Ortaya çıkan grafik tabanlı yapılar çok karmaşıktır. Düğümler arasındaki ilişkiler birçok türde olabilir. Akademik alanda yapılan araştırmalar göstermiştir ki sosyal ağlar ailelerden milletlere kadar pek çok düzeyde kullanılmakta ve problemlerin çözümünün belirlenmesinde, kuruluşların çalışmasında ve bireylerin hedeflerine ulaşmada başarılı olma derecelerinde kritik rol oynamaktadır. En basit haliyle bir sosyal ağ, düğümler arasında arkadaşlık gibi belirli bağların bir haritasıdır. Bireyin bağlı olduğu düğümler o bireyin sosyal temasıdır [36]. Yoğun olarak kullanılmaya başlanan sosyal ağ siteleriyle beraber "sosyal ağ" terimi kişiler arasında yoğun bir şekilde ifade edilmeye başlanmış olsa da, aslında sosyal ağlar internet var olduğundan beri bulunan bir yapıdır. Örnek olarak, kendi içlerinde e-posta ile yazışmakta olan kişilerin meydana getirdiği bir sistemde bulunduracağı çizgisel veri yapısından dolayı sosyal bir ilişkiye bağlı bir ağ olarak ifade edilebilir. Bunun haricinde anlık yazışma programlarındaki (Messenger, Skype vb.) konuşma programları da sosyal ağ olarak belirtilebilir [37]. Sosyal varlıklar arasındaki ilişkileri ve bu ilişki modellerinin uygulamalarına odaklanması, sosyal ağ analizini diğer yöntemlerden ayıran en temel özelliğidir. Sosyal ağ analizi; kişilerin davranışları, inançları ve tutumları yerine, bir sosyal varlık veya aktörler ile etkileşimini ve bu etkileşimin nasıl bir çerçeve veya yapı meydana getirdiğini incelemektedir [38].

4.2.1.Neden Sosyal Ağ Analizi?

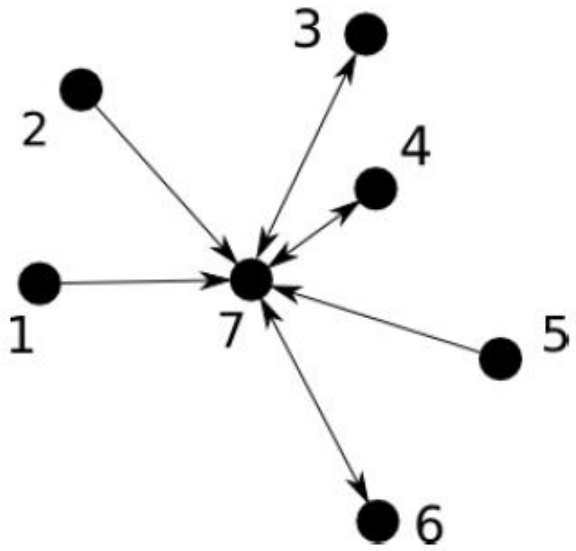
Veri madenciliği, daha önceden kolaylıkla görülemeyen ilişkileri çıkarabilmekteydi. Örneğin; kullanıcıların alışveriş alışkanlıklarına göre birbirine yakın olduğu zaman daha çok satılabilecek ürünlerin birbirlerine yakın konulması ve bu şekildeki ilişkilerin çıkarılması durumunda sağlanacak kazanıma güzel bir örnektir. Benzer şekilde sosyal ağlardaki yapı da buna benzer kazanımlar verebilmektedir. Sosyal ağlar gerek görsel yöntemlerle gerekse matematiksel yaklaşımlarla analiz edildiğinde oldukça ilginç ve bir o kadar faydalı sonuçlar elde edilebilir. Örneğin bir topluluk davranışı ortaya çıkarılabilir [39].

4.2.2.Sosyal Ağ Analizinin Gelişimi

Sosyal ağ analizi (SAA) günümüze kadar pek çok aşamadan geçmiştir. Wolfgang Köhler'den etkilenen Alman Göçmenler, 1930 yılında kavramsal ve sosyal psikoloji üzerinde çalıştılar ve bu çalışmalar, grup dinamikleri üzerinde çalışmalara oldukça faydalı oldu. Laboratuar teknikleri yardımıyla sosyal gruplar arasındaki bilgi akışını incelediler.

1937'de Moreno, sosyogram'ı buldu. Sosyogram; bireylerin, nokta ile bireyler arasındaki ilişkilerin bağlarla gösterimiydi. Bu her ne kadar günümüzde çok büyük bir fikir olarak algılanmasa da 1930'larda hiç kimse toplumun yapısını analitik olarak göstermeyi bulamamıştır.

Moreno'ya göre sosyogram incelendiğinde gruptaki liderler tarafından, ilişkinin akışı gibi bilgiler ortaya çıkarılabilir. Sociometrik yaklaşım bir gruptaki lider ya da star yapısı Şekil 4.5'de gösterilmiştir.

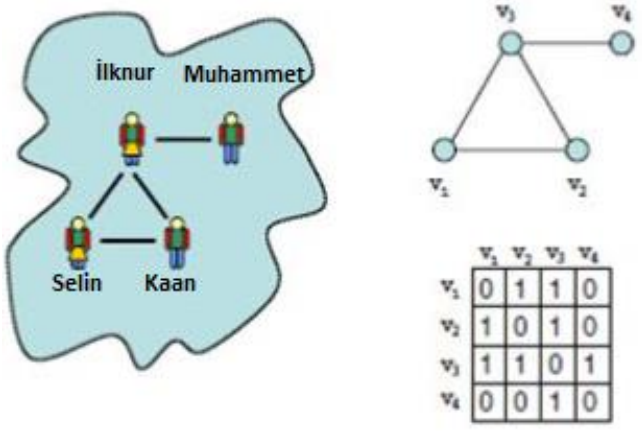


Şekil 7 Sociometrik star

Şekil 4.5.'deki sosyogramda 1,2,3,4,5,6 numaralı noktalar 7 numaralı nokta üzerinde birbirleriyle bağlıdırlar. Bunun yanında 7 numaralı nokta sahip olabileceği maksimum bağlantı sayısına (n-1: n nokta sayısı) sahiptir. Görüldüğü gibi bazı noktalardan 7'ye tek yönlü bağlantı varken bazılarında ise durum çift yönlüdür [39].

4.2.3.Karmaşık Ağların Yapıları

Karmaşık ağ yapısının gösterilmesi için farklı iki yöntem kullanılmaktadır. Biri komşuluk listesi (adjacency list) yöntemidir. Bu yöntem, tüm düğümlerin bitişik veya komşu bulunduğu her düğüm ile bir bütün halinde gösterilmesidir. Bir düğüm, komşuluk listesi kontrol edilerek bulunur. Diğer ikinci yöntemse çok tercih edilen komşuluk matrisi (adjacency matrix) yöntemidir. Bu yöntem ile ise graf içindeki n tane düğüm için meydana getirilen n*n büyüklüğündeki matris içerisinde birbirlerine komşuluğu olan düğümlerin girdileri 1, komşuluğu olmayan düğüm girdileriyse 0 ile gösterilir. Şekil 4.6.'da sosyal ağın graf ve bu graftaki düğümler ile kenarlardan oluşan komşuluk matrisine örnek bulunmaktadır. Matris değerleri girilirken de tüm girdi değerlerine bakılarak, karşılık gelen sütun ve satırdaki düğümlerin içerisinde kenar olup olmadığı belirtilebilir [40]. Karmaşık ağ analizi yöntemleri, çoğunlukla graf modelini matrislere çevirerek çalışmaktadır. Karmaşık ağ analiz ölçütlerinin bulunması adına geliştirilmiş algoritmalar matris tabanlıdır [41].



Şekil 8 Sosyal ağın graf ve matris gösterimi [41]

4.2.4.Karmaşık Ağ Analizdeki Ölçütler

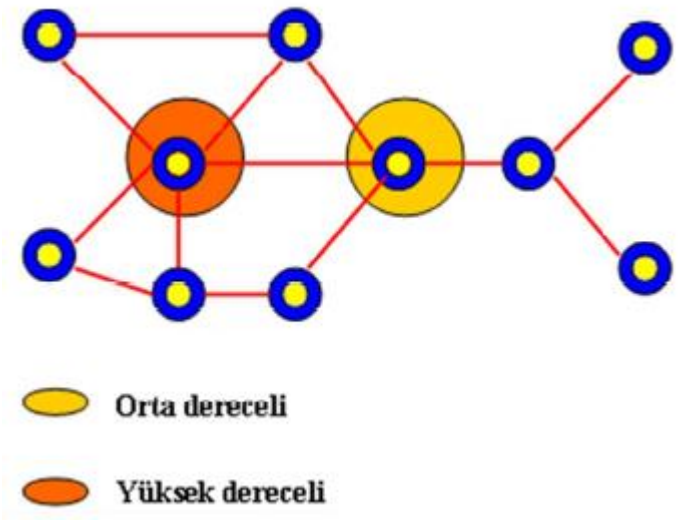
Karmaşık ağ analizi, ağ yapılarının kontrol edilmesi sırasında ağ da bulunan aktörlerin özelliklerinden çok, aralarındaki bağlantıyı incelemektedir. Analiz yöntemi esnasında ağ içerisindeki aktörler (kişiler, bilgisayarlar, kuruluşlar, vb.) aralarındaki bağlantılar kontrol edilerek, sayısallaştırılmasıyla ortaya çıkacak sonuçlara ölçüt denilmektedir. Karmaşık ağ analizinin değişik düzeylerde yapılmasıyla değişik ölçütler alınabilir. Ağın geneli, ağdaki alt gruplar ve ağdaki tüm aktörler için farklı analiz sonuçları ortaya çıkmaktadır. Bunlara göre sosyal ağ analizi ölçütleri bir alt başlıkta incelenmiştir [42].

4.3.Ağ Seviyesinde Ölçütler

Sosyal ağ bağlantılarının ve tüm aktörlerin olduğu, yani ağın bütünü için yapılan analizlerde birtakım ölçütlere ulaşılabilir. Bu ölçütler içerisinden en yaygını yoğunluk (density) ölçütüdür. Ağ içerisinde bulunan tüm bağlantıların oranı ile olası tüm bağlantıların oranı yoğunluk ölçütünü vermektedir. Öğelerin tümünün birbirlerine bağlı bulunduğu ağlarda yoğunluk değeri 1 çıkacaktır. Yoğunluk değeri büyük ise ağ içerisindeki öğeler birbirlerine daha iyi bağlı bulunduklarını ve birbirinden soyutlanmış şekilde olmadıklarını ifade eder. Yoğunluğun haricinde tüm ağlar için kullanılmakta olan bir başka kriter ise büyüklük (size) kriteridir. Sosyal ağ içindeki aktörlerin sayısı o ağın büyüklüğünü göstermektedir. Bir ağın büyüklüğü aktör sayısının çokluğuna bağladır. Aktör sayısı arttıkça ağın büyüklüğü de artmaktadır. Bunu bir örnek ile belirtmek gerekirse; bir okulda 10 öğrencinin bulunduğu bir sınıf düşünelim. Öğrencilerin sayısı az olduğu için birbirlerine kısa sürede alışacaklardır. Bu yüzden aralarındaki ilişki daha güçlü olacak ve ders notları paylaşımları daha fazla olacaktır.

Bir de öğrenci sayısının 100'den fazla olduğu bir sınıfta öğrencilerin birbirlerini tanıma ihtimali çok az ve çoğu öğrenci arasında bir ilişki dahi olamama durumu mevcuttur. Böyle bir durumda birbirleriyle ders notlarını paylaşma gibi iletişimde bulundukları karmaşık ağı oluşturabilmek oldukça zor olacaktır [42, 43].

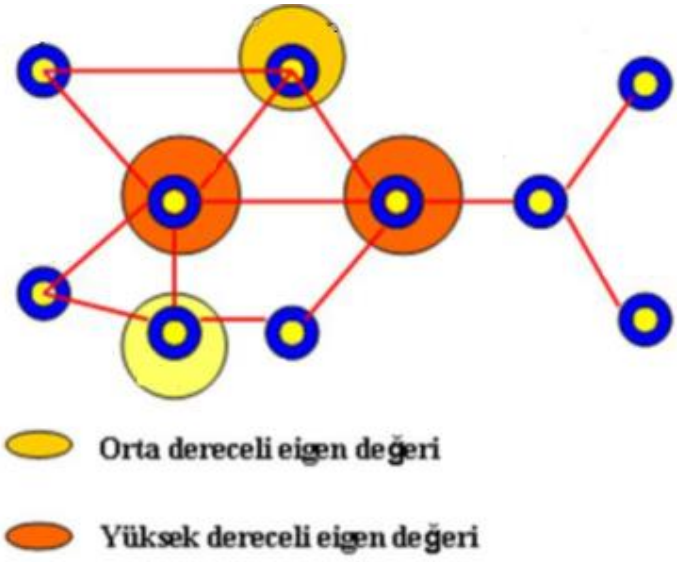
Merkezi ölçütlerden en kolayı derece (degree) denilmekte olan derece merkeziliği (degree centrality) yöntemidir. Ağ içindeki bir aktörün derecesi kendisine bağlı bulunan bağlantıların sayısıyla hesaplanmaktadır. Kolay hesaplanmalar dışında, aktörün önemini gösteren ölçütlerden de birisidir. Bir çok karmaşık ağda, kişinin sahip olduğu ne kadar çok bağlantı varsa o kadar çok önem ve güce sahiptir. Hatta derecesi en büyük olan aktör ağdaki en aktif üye olarak da belirtilebilir. Ağdaki bağlaşların yönlü olması halinde, gelen bağlantı sayıları (in –degree) ile çıkan bağlantı sayıları (out-degree) ile ayrı ayrı hesaplanmaktadır. Şekil 4.7.'de derece merkeziliği yöntemine örnek bulunmaktadır.



Şekil 9 Sosyal Ağda Derece Merkeziliğine Örnek (degree centrality) [44]

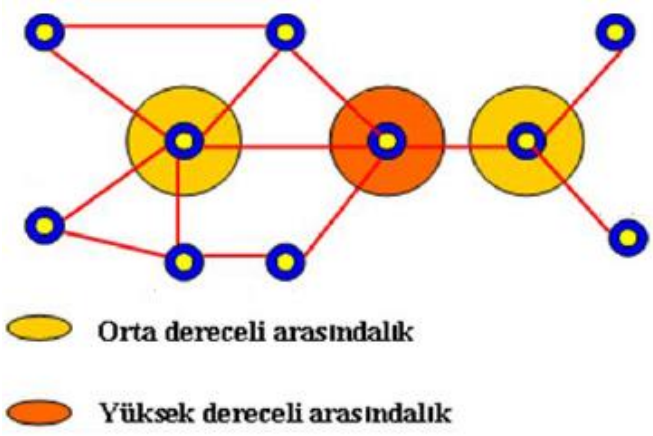
Özvektör merkezilik (Eigenvector Centrality) ölçütü, biraz daha karmaşık derece ölçütü türüdür. Derece hesaplanırken dikkate alınacak aktörün mevcut bağlantılarının eşit olmadığını gösteren ölçüt türüdür. Örneğin; ağ da bulunan bir aktörün önemli bağlantılara olan etkisi önemli olmayan bağlantılardan daha yüksektir. Bağlantısı yüksek olan öğelerin daha merkezde bulunması o öğenin daha merkezi bir noktada bulunduğunu belirtir.

Özvektör merkezilik ölçütü hesaplanırken komşuların merkezilik derecelerinin toplamı dikkate alınmaktadır [44]. Şekil 4.8.'de özvektör merkeziliği ölçütüne örnek ağ bulunmaktadır.



Şekil 10 Sosyal ağda özvektör merkeziliğine örnek (eigenvector centrality) [44]

Diğer bir yöntemse arasındalık merkeziliği (betweenness centrality) yöntemidir. Merkezilik yöntemleri arasında en karışık hesaplanması olanıdır. Ağdaki bir düğüm veya öğe kısayolların oranıyla hesaplanır. İlk olarak ağ içerisindeki bütüm düğüm çiftleri içerisinden kısayollar (geodesics) çıkarılır, sonra bu yolların geçtiği düğümlerin oranı arasındalık ölçütünü çıkarır. Arasındalık ölçütü yüksek bulunan aktörler konumları nedeniyle, diğer öğelere göre daha önemli konumda bulunurlar [35, 41]. Şekil 4.9.'da arasındalık merkeziliği ölçütüne örnek ağ bulunmaktadır.



Şekil 11 Sosyal ağda arasındalık merkeziliği (betweenness centrality) [44]

Yakınlık Merkeziliği (Closeness Centrality), karmaşık ağda olan bir düğümün diğer düğümlere giden en kısa yolların toplamından (geodesic distance) çıkarılır. Yönlü bağlantı olması halinde en kısayolların bulunması kısmında yönlere dikkat edilmesi gerekir. Böyle durumda giriş (in-closeness) ile çıkış (out-closeness) olarak iki farklı yakındalık ölçütü bulunur [35, 45]. Yakınlık ölçütünde küçük düğümler büyük düğümlere göre daha merkezi durumdadırlar.

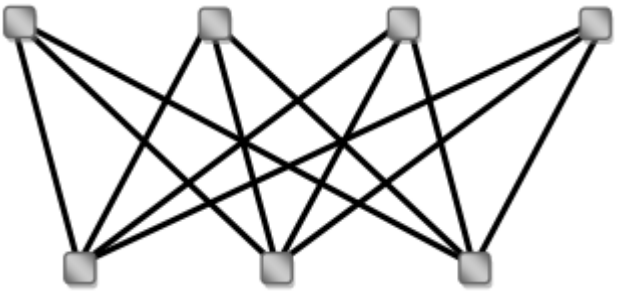
4.3.1. Topluluk Analizi

Karmaşık ağlarda düğümlerin bir araya gelmesiyle belli kümelerin oluşmasına topluluk analizi denir. Ağ içerisinde genellikle iki tür topluluk oluşmaktadır. Bunlardan biri belirgin topluluklar (excplicit community). Bu topluluklarda üyeler gruplara bilinçli şekilde katılır. Ayrıca topluluk içindeki tüm üyeler birbirini bilir. İkinci topluluk ise üstü kapalı topluluklardır (implicit community). Bu topluluklar grup oluşturma amacı olmadan kendi aralarındaki etkileşim sonucunda meydana gelmektedirler [46].

Ağda üstü kapalı toplulukları bulmak amacıyla topluluk tespiti (community detection) yapılmaktadır. Üç farklı yöntem ile topluluk analizi bulunmaktadır. Bunlar; kenar merkeziliği tabanlı, modülarite tabanlı ve benzerlik tabanlı topluluk analizi yöntemleridir [46].

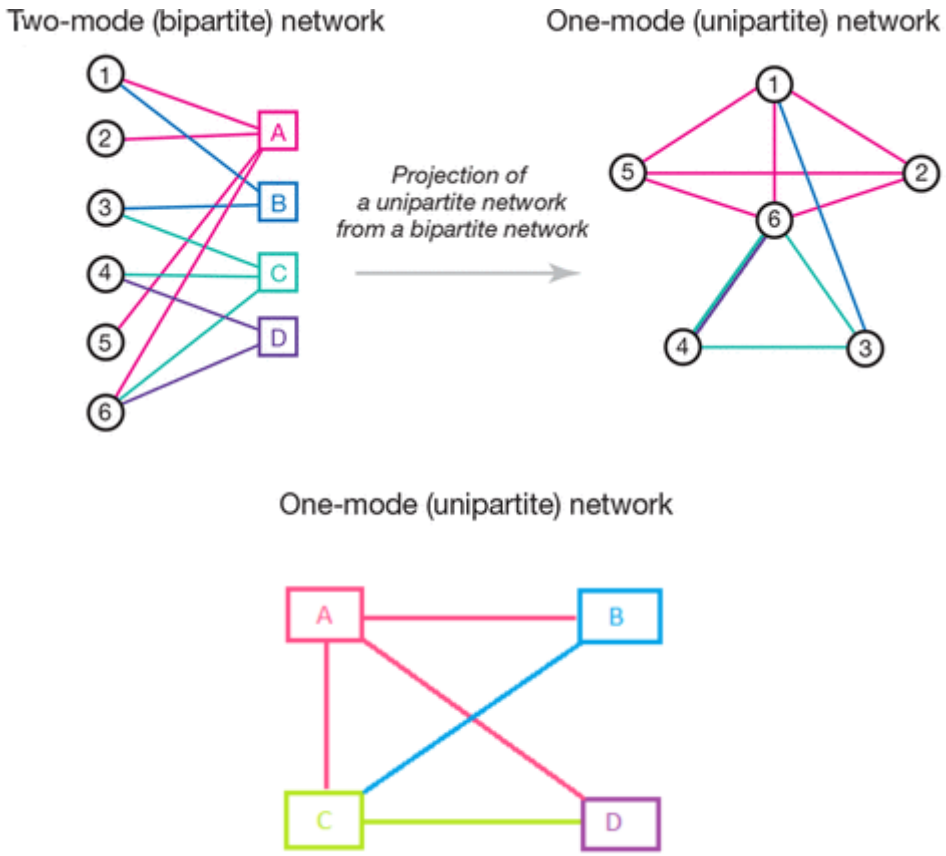
4.3.2. İki Parçalı (Bipartite) Graflar

İki Parçalı Graflar, veri modellemede sıklıkla kullanılan grafların (graph) özel bir durumudur. Eğer grafı oluşturan düğümler iki farklı kümeye ayrılabiliyor ve iki küme içerisindeki düğümlerde birbirine kenarlar (edges) ile bağlanıyorsa iki parçalı graf (bipartite graph) denmekte olup örnek Şekil 4.10.'daki gibidir.



Şekil 12 İki Parçalı Graf (Bipartite Graph)

İki parçalı graf (bipartite graph) üzerinde birbirine kenarlar ile bağlanmış olan kenarlar bağlantı durumlarına göre Şekil 4.11.'deki gibi tek parça (One mode Projection) graf halini almaktadır.



Şekil 13 İki Parçalı Graf → Tek Parçalı Graf

BÖLÜM 5. I. YÖNTEM – BİRLİKTELİK ANALİZİ

5.1.Veriler ve Toplanması

Bu çalışmada kurumsal iki firmanın Aralık 2017 – Aralık 2018 tarihleri arasındaki satınalma verileri incelenmiştir. Firmaların satınalma verileri üzerinde birliktelik kuralları ve sosyal ağ analizi yöntemi kullanılarak iki farklı değerlendirme yapılmıştır. Bu verilerde birinci firmanın 23.447 satır satınalma kaydı bulunmakta olup, ikinci firmanın ise 3.348 satır satınalma kaydı bulunmaktadır. Tablo 5.1.'de birinci firmanın, Tablo 5.2.'de ise ikinci firmanın satınalma verileri bulunmaktadır.

	A	B	C
1	**Sipariş No**	**Madde Kodu**	**Madde Adı**
2	201800001	YMFST000127	FS8079|H14|0,21|1420|DZ|Ambalaj
3	201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
4	201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
5	201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
6	201800001	YMRLO001777	RL3105|H26|0.70|1285|DZ|B_kompnl
7	201800012	AH00100001	ALUMİNYUM HAMMADDE; %99,7; (Alaşımsız)
8	201800012	AH00200001	ALUMİNYUM HAMMADDE; %99,6; (Alaşımsız)
9	201800012	AH00200002	ALUMİNYUM HAMMADDE; %99,5; (Alaşımsız)
10	201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
11	201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
12	201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
13	201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
14	201800020	SM00100003	SHELL; Ø1128X2350; T-DH3/4/5; ; R/N:1143003-503-04
15	201800020	SM00100003	SHELL; Ø1128X2350; T-DH3/4/5; ; R/N:1143003-503-04
16	201800020	SM00100004	SHELL; Ø1128X2540; T-DH6/7; ; R/N:1163003-503-04
17	201800020	SM00100004	SHELL; Ø1128X2540; T-DH6/7; ; R/N:1163003-503-04
18	201800020	SM00100006	SHELL; Ø880XØ1025 XØ1940; T-DH2; ; R/N:1123003-503
19	201800020	SM00100006	SHELL; Ø880XØ1025 XØ1940; T-DH2; ; R/N:1123003-503
20	201800039	AALTD0005	TL5754|H22|2|1250|DZ|TicLevHR
21	201800039	AALTD0006	TL5754|H22|2|1500|DZ|TicLevHR
22	201800039	AALTD0008	TL5754|H22|3|1250|DZ|TicLevHR
23	201800039	AALTD0062	TL5754|H22|1|1000|DZ|TicLevHR
24	201800039	AALTD0088	TL5754|H22|3|1500|DZ|TicLevHR
25	201800040	MM00100087	MASURA; KARTON; Ø545 MM X Ø505 X 1530 MM; 5100N;
26	201800040	MM00100088	MASURA; KARTON; Ø545 MM X Ø505 X 1530 MM; 6300N;

Tablo 5.1. Tez Çalışmasına Konu Olan Birinci Firmanın Veri Örneği

Sipariş No	Madde Kodu	Madde Adı
SA00113921	431313	ETİKET 2 LT BİDON MEXICANA SOS/ COLORADO (Ö.A)
SA00113921	432161	ETİKET 300 CC MEXICAINE MEDIUM SOS/ LEADER PRICE
SA00113921	432169	ETİKET 900 ML ÇİĞKÖFTE SOS TERS SARIM/ KOMAGENE (Ö.A)
SA00113921	433024	ETİKET 20 GR ACI SOS (820169)/ COLORADO
SA00113930	353008	KURUTULMUŞ NANE YAPRAK
SA00113930	353019	KURUTULMUŞ KEKİK YAPRAK
SA00113938	493128	ÜST FOLYO TAHİNLİ YOĞURTLU SOS / KFC
SA00113938	493130	STICK ACI KETÇAP 1080 MM/ KINGTOM
SA00113956	502068	SPİRALLİ AJANDA 17X24 CM
SA00113956	502068	SPİRALLİ AJANDA 17X24 CM
SA00113956	502069	KÜÇÜK AJANDA 13X21 CM
SA00113956	502070	MASA TAKVİMİ 14,8 X 21 CM
SA00113962	OTDE0032	(500052771)Avansas Clean Çamaşır Suyu 5 kg
SA00113962	OTSA0043	(500070130)BEYBİ NİTRİL MUAYENE ELDİVEN MAVİ MEDİUM 100LÜ
SA00113962	OTSA0064	(500063433)Avansas Mikrofiber Temizlik Bezi 5'li Paket
SA00113974	503035	OGYE AGAR BESİYERİ/MERCK 1.05978
SA00113974	503036	OGYE AGAR KATKISI/MERCK 1.09877
SA00113974	503038	REINFORCED CLOSTRIDIAL MEDIUM/OXOID CM 0149
SA00113974	503039	BUFFERED PEPTON WATER/OXOID CM 0509
SA00113974	503044	HY-LITE TEST KİTİ/MERCK
SA00113974	503045	MRS BOTH/MERCK 1.10661
SA00113974	503152	STERİL PLASTİK PET PP NUMUNE ŞİŞESİ/250 ML

Tablo 5.2. Tez Çalışmasına Konu Olan İkinci Firmanın Veri Örneği

5.2.KNIME Uygulaması ve Detayları

Konstanz Information Miner (KNIME), Konstanz Üniversitesinin görsel veri madenciliği araştırma ekibi tarafınca Eclipse Rich Client Platform içerisinde bulunan bir yazılım türüdür. KNIME, geliştirmeye açık özellikleriyle öne çıkmaktadır. Kullanıcılarına yazılım geliştirme eki vererek, kişilerin kendileri için modül getiştirebilmelerini sağlamakta olan tek yazılımdır [47]. KNIME (Konstanz Information Miner) içerisinde bulundurduğu düğümlerle beraber (node), değişik durumlarda çalışabilecek Java tabanlı veri madenciliği yazılımıdır [48].

Açık kaynak kodlu bir veri analizi platformudur. İçerisinde büyük veri ile ilgili iki araç bulunmaktadır. KNIME Big Data Connector aracı ile KNIME Server ve KNIME

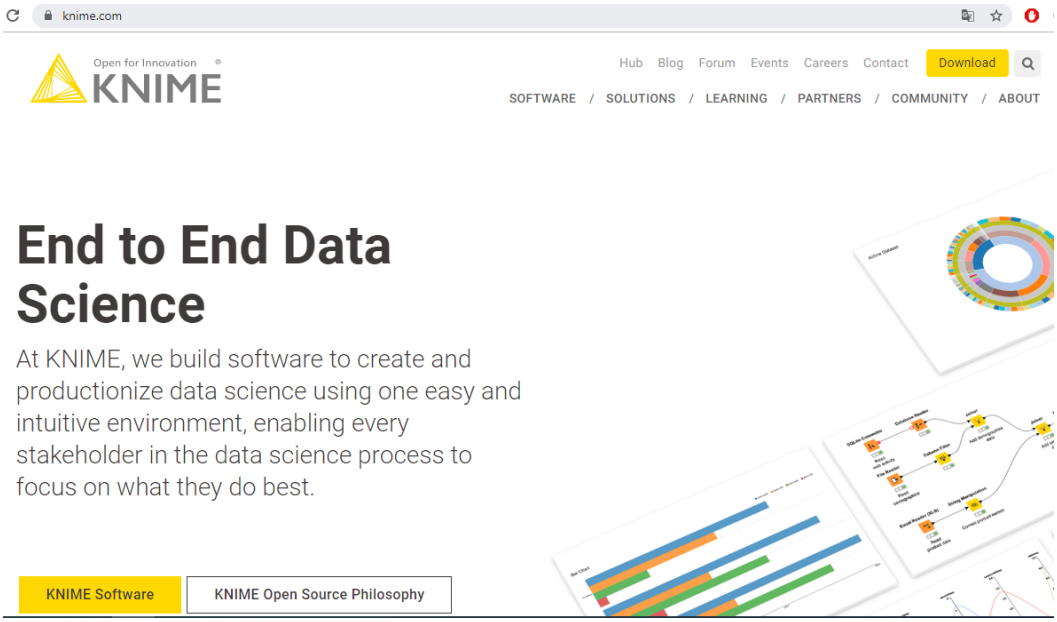
Analytics içerisinden Hadoop verilerine kolayca erişim imkanı sunmaktadır. Hadoop/HDFS'ye erişmek için Hive veya Impala üzerinden kütüphaneler aracılığıyla ulaşılır. Ayrıca bu verilere eriştikten sonra KNIME düğümlerine veriler eklenerek işlemler yapılır. KNIME platformu birçok işlemleri gerçekleştirmektedir. Bunlar; belge okuma ve ayrıştırma, varlık tanıma, filtreleme ve değiştirme işlemleri, kelime sayma, anahtar kelime çıkarımı, dönüşüm ve görselleştirme işlemleri olarak belirtilebilir [49].

KNIME, verilerde saklı olan potansiyelin keşfi, ve yeni öngörüler elde etmek veya yeni oluşumları tahmin etmek için tasarlanmış bir uygulamadır. Birçok sektörde, yıllardan beri 60'tan fazla ülkede farklı veriler ve platformlar üzerinde çalışmalar yapılmaktadır. Örneğin; rakamlardan görüntülere, moleküllerden insanlara, sinyallerden karmaşık ağlara ve büyük veri analizine kadar birçok alanda KNIME uygulaması kullanılmaktadır [8].

KNIME uygulaması https://www.knime.com adresi üzerinden ücretsiz olarak indirilebilir. Sitede “Learning” bağlantısı içerisinde KNIME kullanımı hakkında genel bilgiler ve eğitim dokümanları yer almaktadır. “Software” bağlantısı altında yazılım ile ilgili bir takım genel bilgiler yer almaktadır.

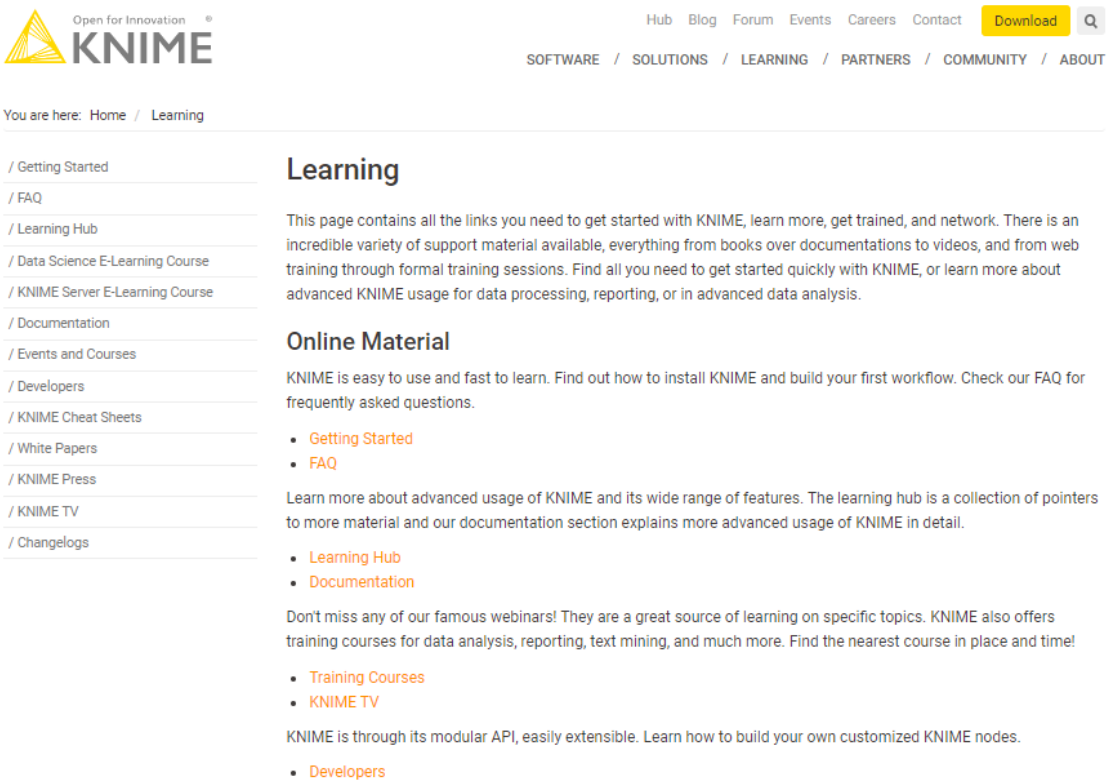
5.3.KNIME İndirilmesi ve Kurulumu

Bu bölümde KNIME’ın nasıl indirileceği ve bilgisayara kurulumu açıklanacaktır. Knime.com web sitesinden software sekmesine girerek indirilebilir.



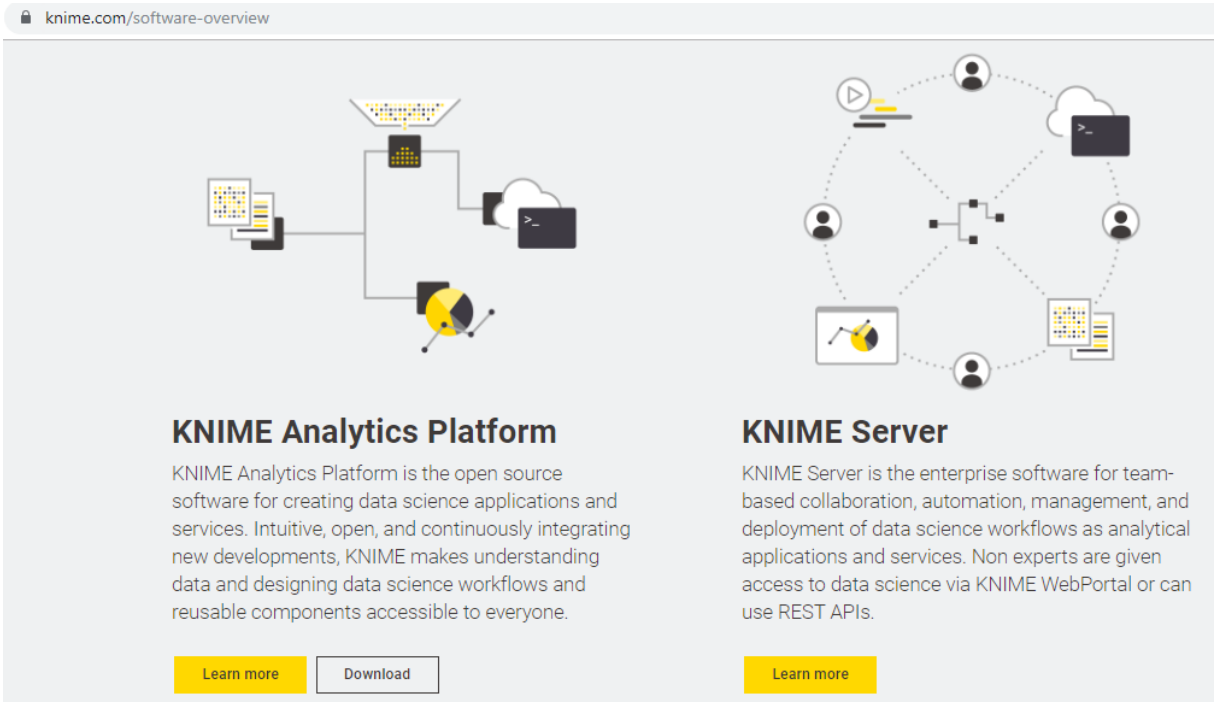
Şekil 14 KNIME Ana Sayfa

Şekil 5.1.'de görüldüğü gibi sitede genel seçenek olarak sotfware, solutions, learning, partners, community ve about vardır. Konferanslar, duyurular, yarışmalar gibi etkinliklerin bilgileri bu web sitesinde yayınlanmaktadır.



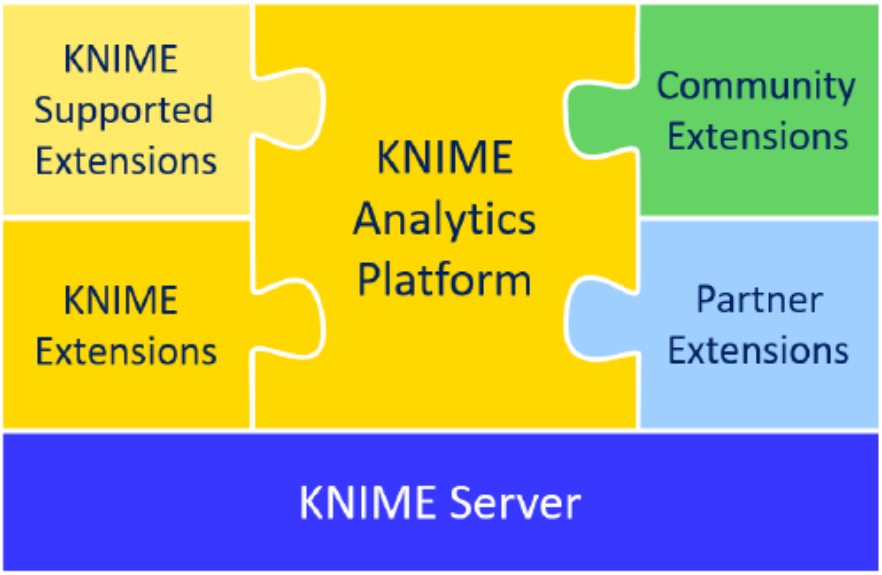
Şekil 15 KNIME Eğitim Serileri

Şekil 5.2.'de en üst kısımda görülen learning başlığı altındaki eğitim serileri bulunmaktadır.



Şekil 16 KNIME Software Sayfası

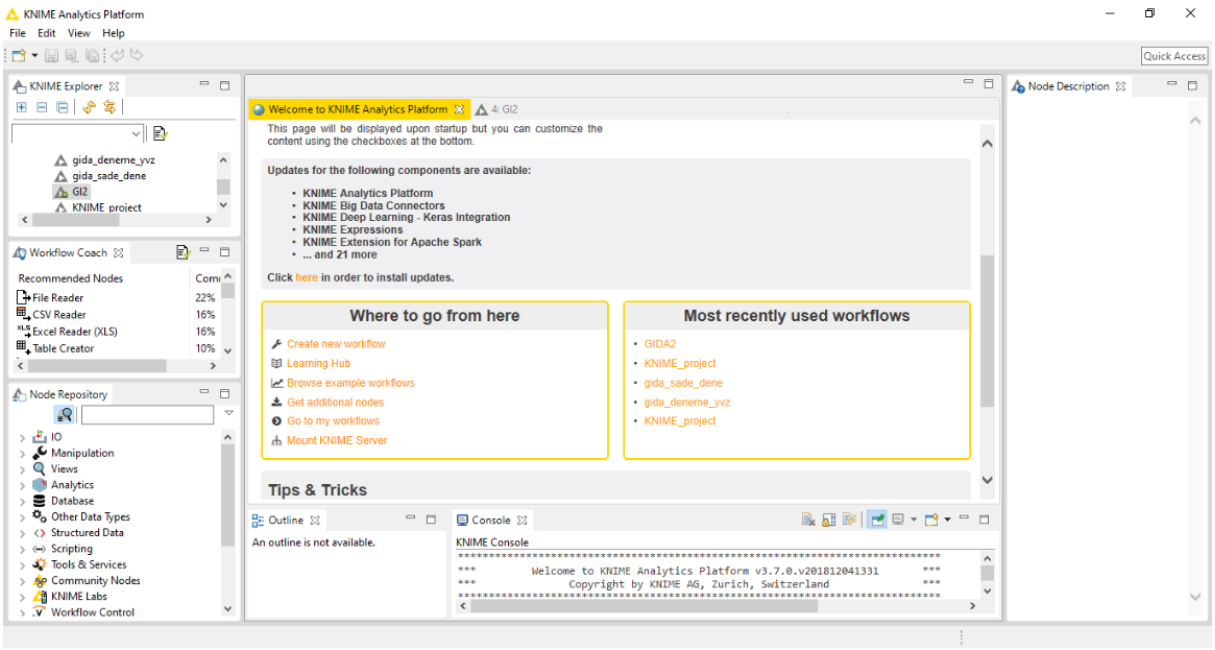
Resim 5.3.'de görülen software'in altında yazılımla ilgili genel bilgiler bulunmaktadır. Data mining suitlerinin tek bilgisayar üzerinde çalışamayacak kadar büyük olmasından dolayı artık KNIME server da görülmektedir. SAS, Rapid miner gibi diğer tool'larında bu yüzden kendi sunucuları bulunmaktadır. Bu çalışmada, Knime'ın analytics Platform'u kullanılacaktır. KNIME yazılımına gelen open source extansions'lar bulunmakta ve onun dışında değişik partnerler veya community (topluluk) tarafından üretilen extansions'larda bulunmaktadır. Anlamı bazı extansions'ların open source yani ücretsiz olarak kullanılabilirken bazılarının ücretli olmasıdır. KNIME analytics platform, 20000 modül ve yüzlerce örnek içeren oldukça kapsamlı bir program olup bu çalışma için KNIME alaytics platform kullanılacaktır. Resim 5.4.'de KNIME uygulamasının genel yapısı bulunmaktadır.



Şekil 17 KNIME Uygulamasının Yapısı

KNIME uygulamasının işletim sistemine göre Windows, Linux, Mac, SDK (Software Development Kit) seçenekleri mevcuttur.

5.4.KNIME Uygulamasının Çalıştırılması

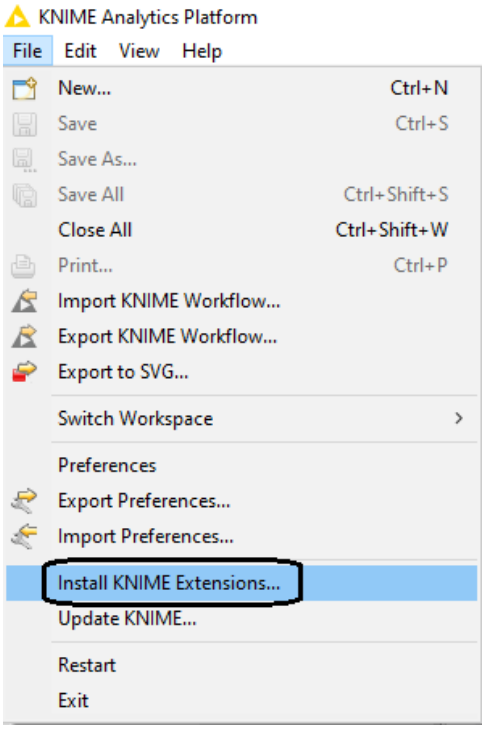


Şekil 18 Programın ilk açılıştaki görüntüsü.

Şekil 5.5.'de KNIME programının ilk açılış ekranı bulunmaktadır. Orta alan, projenin yaratılacağı ve verilerin alınıp analiz edileceği kısımdır. Sağ alt köşede bulunan Node Repository isimli pencerede ise program ile alakalı her türlü modele, algortimaya ve uygulamaya erişilebilmektedir. Sağ üst köşede ise Node Description pencerisi, seçtiğimiz öğeyle alakalı ufak açıklamaları kullanıcı tarafına sunmaktadır. Sağ alt köşede bulunan Console penceresi çalıştırdığımız algoritma ya da program parçacığının, yaptığını, varsa hatasını ve nedenini kullanıcıya gösterir.

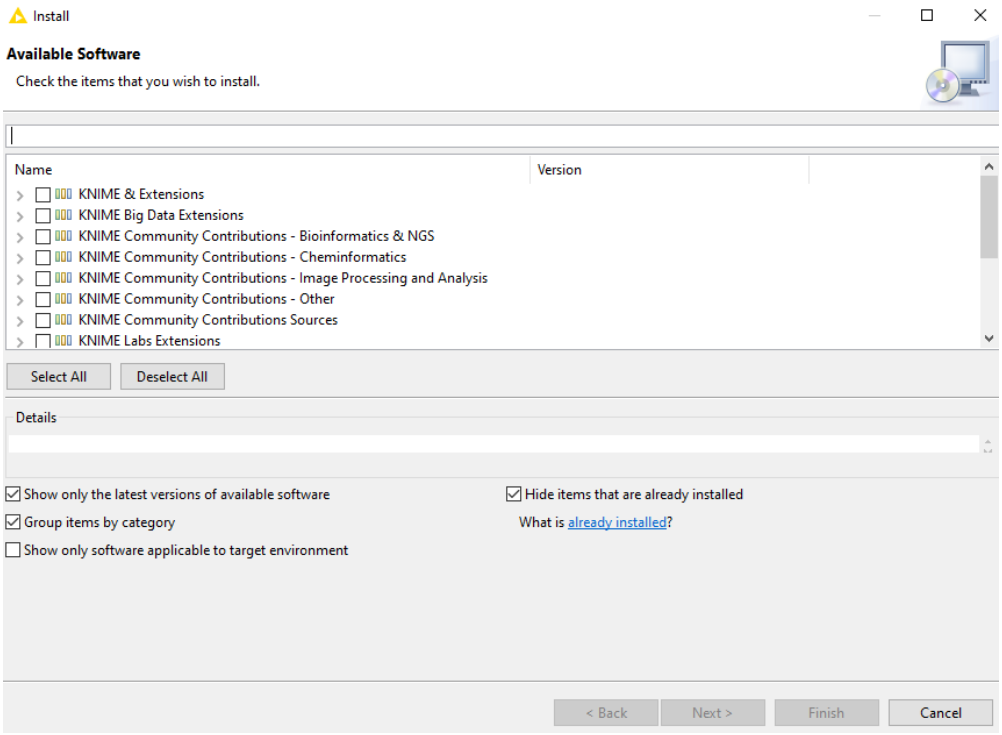
KNIME açıldıktan sonra görülen ilk ekrandır ve bazı alt pencerelerden oluşmaktadır. KNIME explorer çalışmaların kaydedildiği, örneklerin olduğu bölümdür. Workflow coach, çalışan kişiye KNIME'ın tavsiyeler verdiği penceredir. En önemli pencere repository penceresidir. Burada yüklü bazı düğümler vardır. Outline, yukarı pencerede çizilecek olanların ufak görüntüsünün gösterildiği penceredir. Node description ise kullanılan node'un tanımının yapıldığı penceredir.

KNIME menüleri içerisindeki File/Install KNIME Extensions tıklanarak Şekil 5.6.'daki gibi yeni eklentiler eklenebilir.



Şekil 19 KNIME Analytics Plarform – File Menü Ekranı

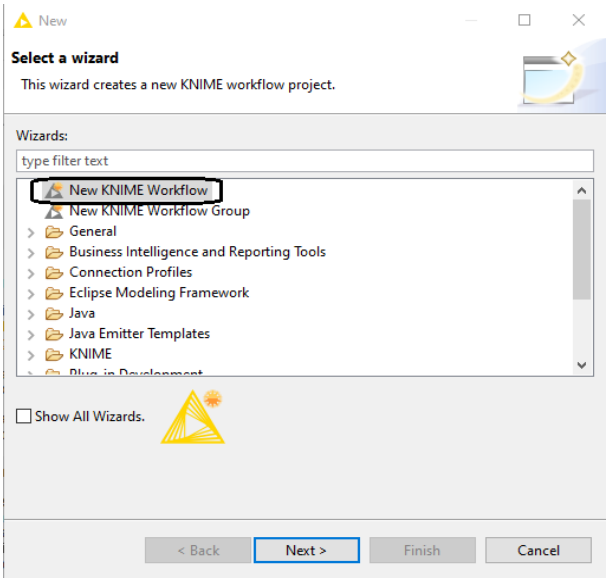
İnternet üzerinden KNIME özellikleri indirmek istediğinizde, Şekil 5.7.'de bulunan ekranda eğer indireceğiniz öğeleri tanımıyorsanız, indirdikleriniz, var olan kendi içindeki öğelerle çakışabilir. Bu durumda yükleme gerçekleşmeyecektir. Örneğin, tüm seçenekleri seçerek yükleme yapmak isterseniz, bu uyarıyla karşılaşabilirsiniz. İndirdiğiniz KNIME versiyonuna bağlı olarak bazı öğeleri yükleyememe şansınız var. Bu aşamada deneme yanılma ile öğe seçebileceğiniz gibi, öneri olarak en üstteki KNIME & Extensions öğelerinin tamamını indirmeyi deneyin. İşletim sisteminiz ve versiyonununuz izin veriyorsa sorun yok demektir. Çakışma olursa bu durumda öncelikle Chemistry özelliklerinden vazgeçerek tekrar deneyebilirsiniz.



Şekil 20 Install KNIME Extensions ekranı

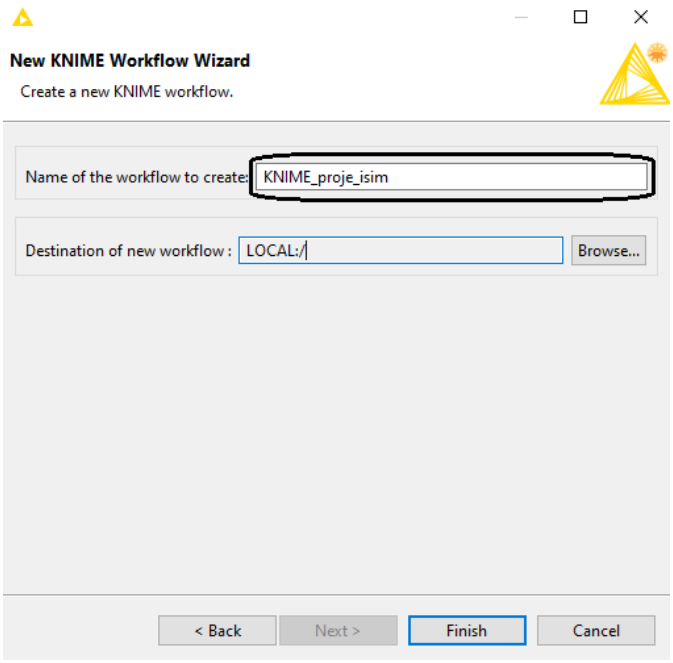
5.5.Çalışma Alanı Oluşturma

KNIME içerisinde çalışma alanı oluşturmak için menülerden File\New seçeneğinden Şekil 5.8 ekranındaki gibi yeni bir workflow açılıp KNIME uygulamasında bir çalışma alanı oluşturulabilir.



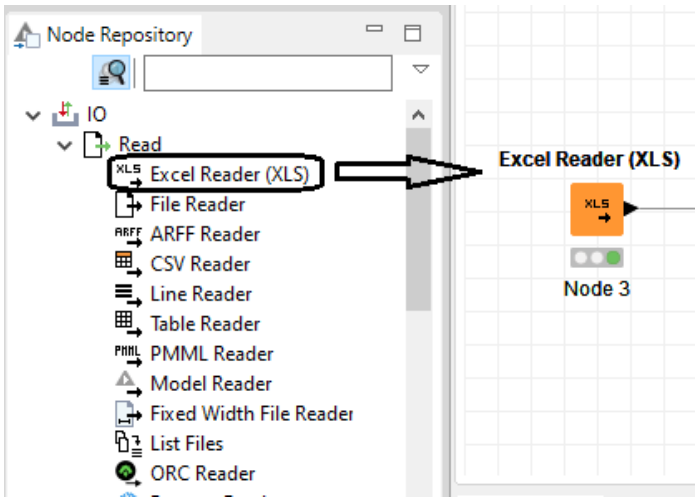
Şekil 21 Yeni çalışma alanı (Workflow) Açma

"New KNIME Workflow" seçildikten sonra açılan workflow'a Şekil 5.9.'daki gibi yeni isim vererek çalışma alanı oluşturulur.



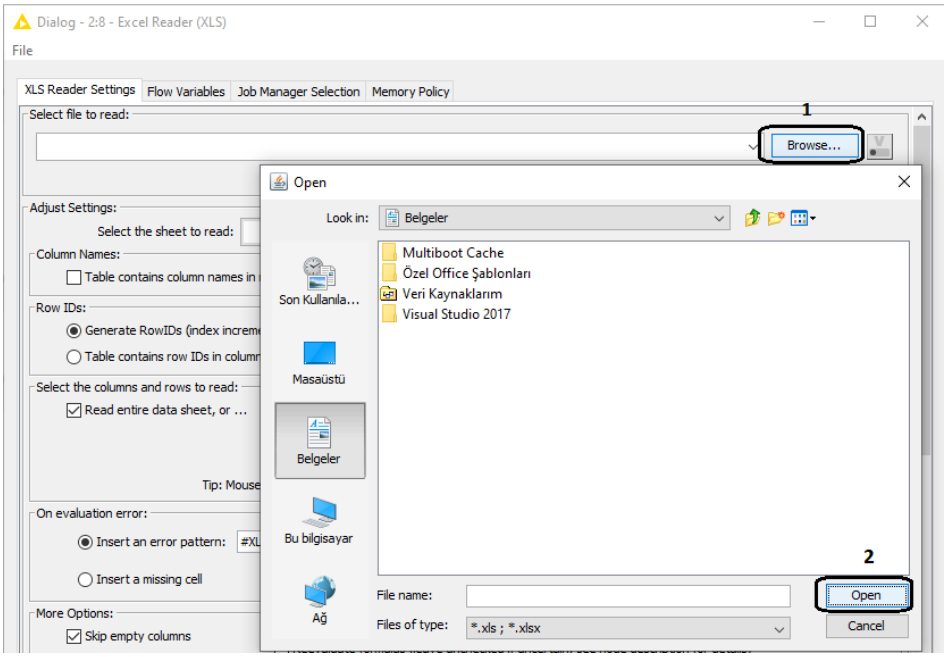
Şekil 22 Çalışma Alanı Oluşturma Sırasında Projeye İsim Verme

Kullanılacak olan veri Excel formatında olduğu için "Node Repository" alanından IO\Read\Excel Reader (XLS) node'u sürükleyerek workflow içerisine Şekil 5.10'daki gibi eklenmelidir.



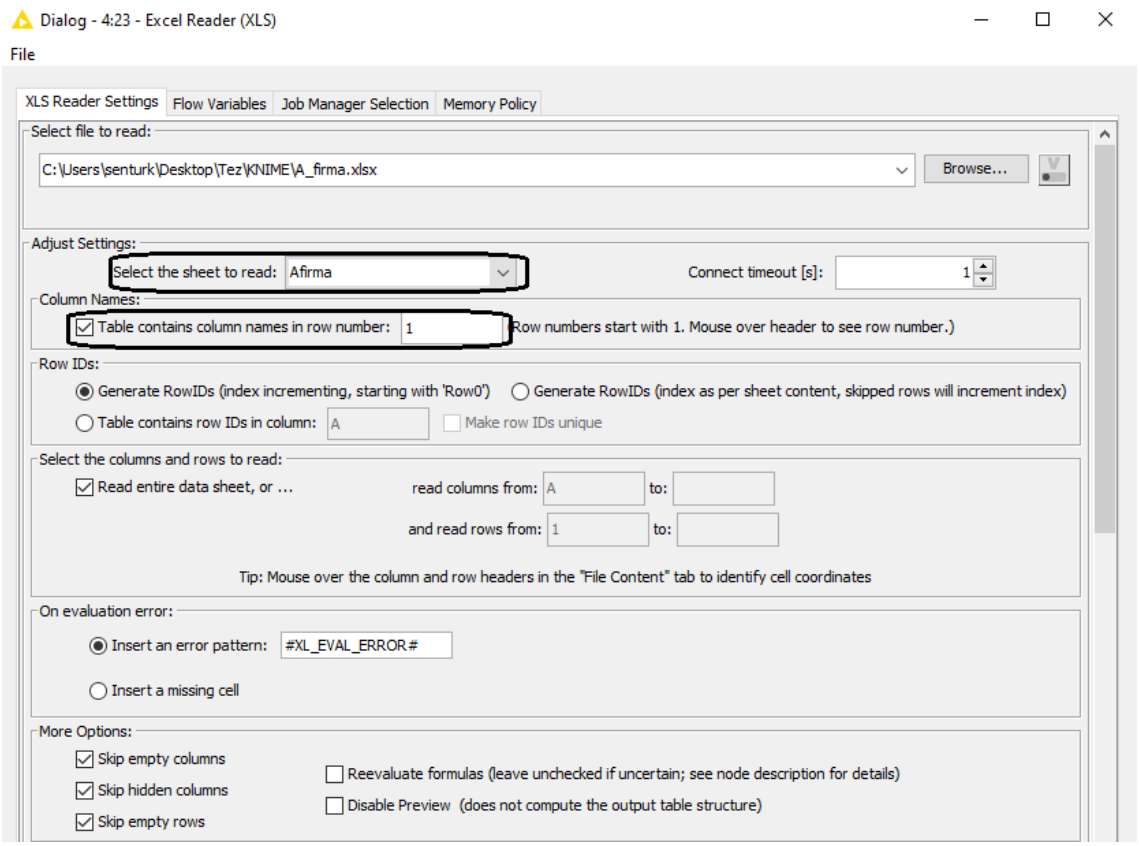
Şekil 23 KNIME Excel Reader (XLS)

Excel Reader, çalışma alanına eklenmesi sonrası çalışma yapılacak veri (Excel dosyası) eklenmelidir. Mouse ile Excel Reader üzerinde sağ tuş yapılması sonrası "Configure" seçeneği tıklanarak istenen veri Şekil 5.11.'deki gibi eklenmelidir.



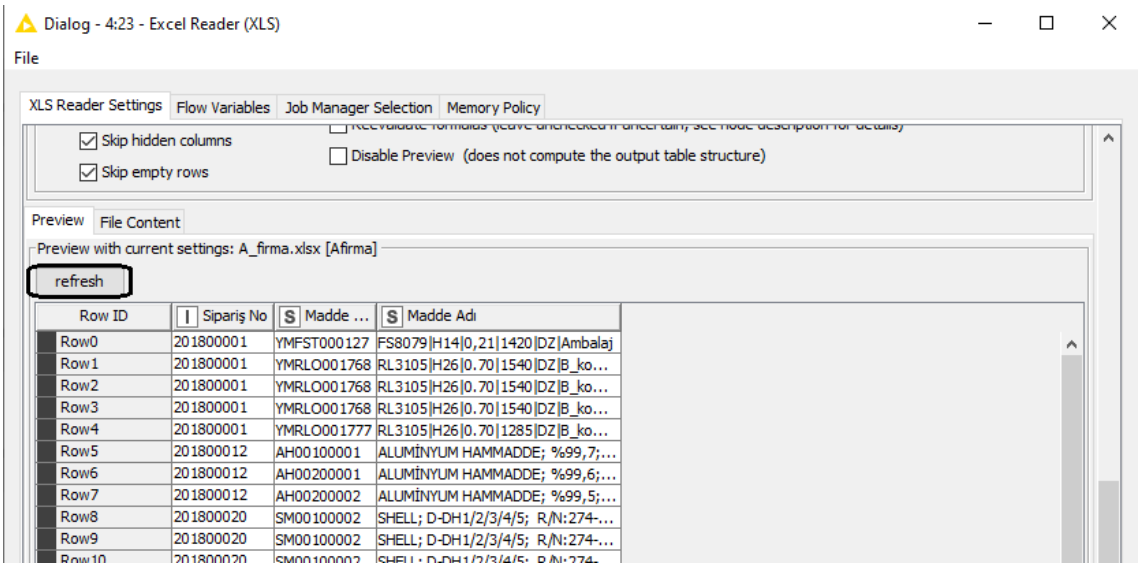
Şekil 24 KNIME XLS Dosyasının Eklenmesi

İstenen veri eklenmesi sonrası Şekil 5.12.'daki gibi Excel dosyası içerisinden çalışılmak istenen sayfa "Select the sheet to read" sekmesinden seçilebilir. Excel sayfasındaki ilk satırdaki verileri başlık olarak kullanabilmek için "Table contains column names in row number" işaretlenmelidir. Excel dosyasındaki ilk satırdaki verileri başlık olarak kullanabilmek için "Table contains column names in row number" seçeneği işaretlenmelidir.



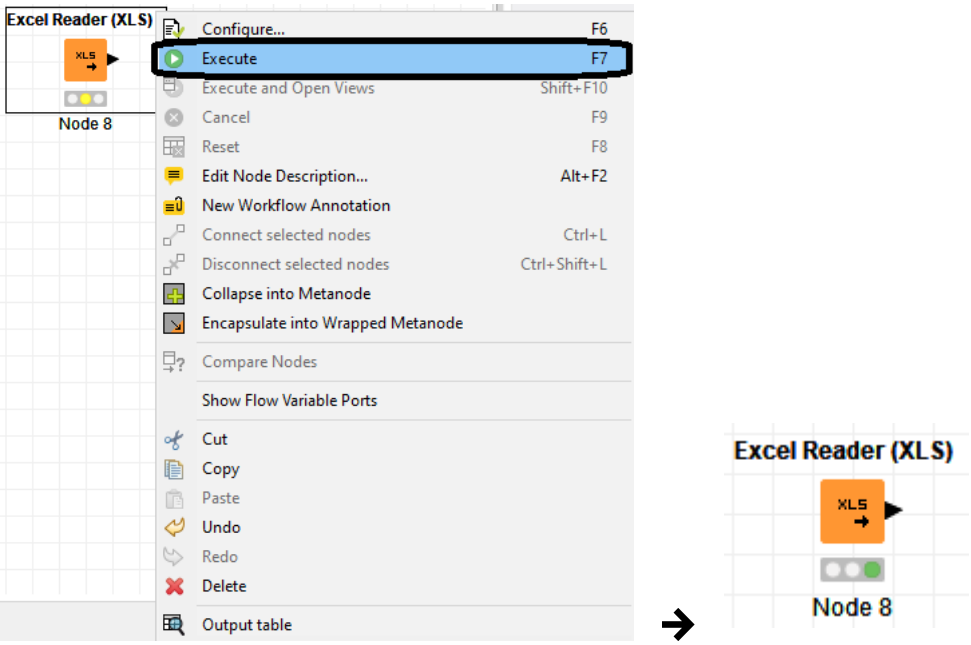
Şekil 25 KNIME Excel Reader Node Configure

Configure alanında herhangi bir değişiklik yapıldığında değişiklikleri görebilmek için Şekil 5.13.'deki gibi sayfanın altında bulunan refresh tuşuna basılmalıdır.



Şekil 26 KNIME Excel Reader Node Configure Refresh

Configure içerisindeki ayarları tamamlandıktan sonra Excel Reader node üzerinde sağ tıklayıp Execute yapıldığında ve F7 tuşu ile Şekil 5.14.'deki gibi "Sarı" durumdan "Yeşil" duruma geçmiş olacaktır. Böylelikle veri bir sonraki node'a aktarılabilir. Node sarı veya kırmızı olması halinde bir sonraki node'a veri akışı olmamaktadır. Node kırmızı olduğunda yapılmış olan Configure ayarlarında hata olduğunu belirtir. Böyle durumda Configure ayarlarının yeniden kontrol edilmesi ve ayarlanması gerekmektedir.



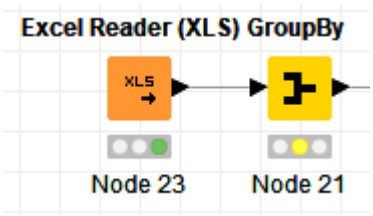
Şekil 27 KNIME Excel Reader Configure ayarları sonrası çalıştırılması

Excel Reader'dan sonra eklenen node GroupBy node'udur. Veriye birliktelik kuralı uygulayabilmek için gerekli formata GroupBy node ile getirildiğinden çok önemli node'dur. Bu node'un configurasyonu içerisinde dikkat etmemiz gereken 2 önemli unsur bulunur. Kullanılan verilerde eğer aynı sipariş numarası içerisinde birden fazla madde kodu var ise aldığı her madde kodu ayrı ayrı satırlarda gösterilmelidir. Bu verilerde her sipariş numarası tek satır, madde kodları ise aralarında virgül olacak şekilde yanyana düzenlenmesi gerekecektir.

GroupBy Node'u ile tekilleştirilecek sipariş no örneği Tablo 5.3.'deki gibidir;

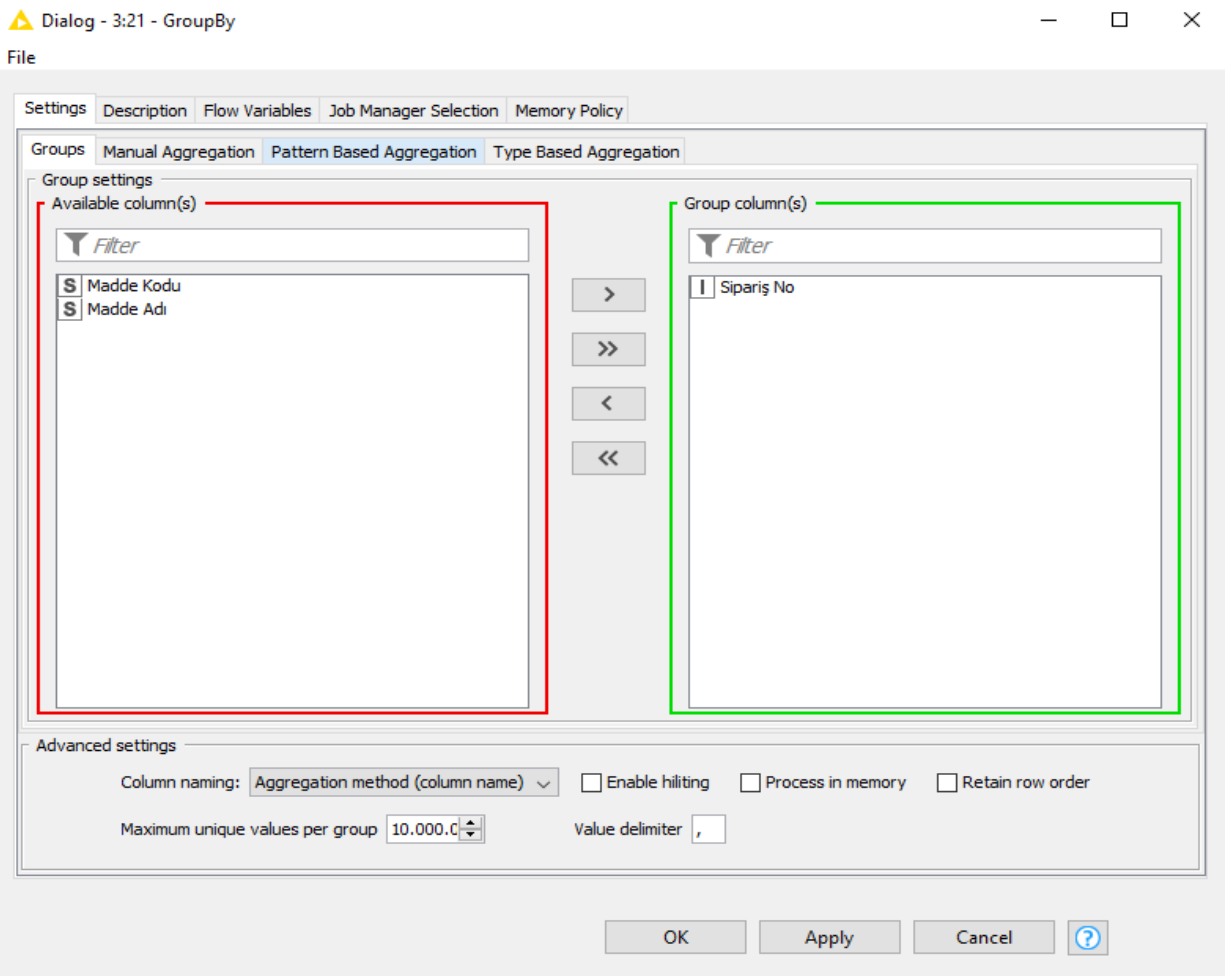
Tablo 5.3. GroupBy Node'u İle Tekilleştirilecek Sipariş No Örneği

Sipariş No	Madde Kodu	Madde Adı
201800001	YMFST000127	FS8079|H14|0,21|1420|DZ|Ambalaj
201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
201800001	YMRLO001768	RL3105|H26|0.70|1540|DZ|B_kompnl
201800001	YMRLO001777	RL3105|H26|0.70|1285|DZ|B_kompnl
201800012	AH00100001	ALUMİNYUM HAMMADDE; %99,7; (Alaşımsız)
201800012	AH00200001	ALUMİNYUM HAMMADDE; %99,6; (Alaşımsız)
201800012	AH00200002	ALUMİNYUM HAMMADDE; %99,5; (Alaşımsız)
201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
201800020	SM00100002	SHELL; D-DH1/2/3/4/5; R/N:274-503-4
201800020	SM00100003	SHELL; Ø1128X2350; T-DH3/4/5; ; R/N:1143003-503-04
201800020	SM00100003	SHELL; Ø1128X2350; T-DH3/4/5; ; R/N:1143003-503-04
201800020	SM00100004	SHELL; Ø1128X2540; T-DH6/7; ; R/N:1163003-503-04
201800020	SM00100004	SHELL; Ø1128X2540; T-DH6/7; ; R/N:1163003-503-04
201800020	SM00100006	SHELL; Ø880XØ1025 XØ1940; T-DH2; ; R/N:1123003-503
201800020	SM00100006	SHELL; Ø880XØ1025 XØ1940; T-DH2; ; R/N:1123003-503
201800039	AALTD0005	TL5754|H22|2|1250|DZ|TicLevHR
201800039	AALTD0006	TL5754|H22|2|1500|DZ|TicLevHR
201800039	AALTD0008	TL5754|H22|3|1250|DZ|TicLevHR
201800039	AALTD0062	TL5754|H22|1|1000|DZ|TicLevHR
201800039	AALTD0088	TL5754|H22|3|1500|DZ|TicLevHR
201800040	MM00100087	MASURA; KARTON; Ø545 MM X Ø505 X 1530 MM; 5100N;
201800040	MM00100088	MASURA; KARTON; Ø545 MM X Ø505 X 1530 MM; 6300N;



Şekil 28 KNIME GroupBy Node Configure ayarları yapılmadan önce

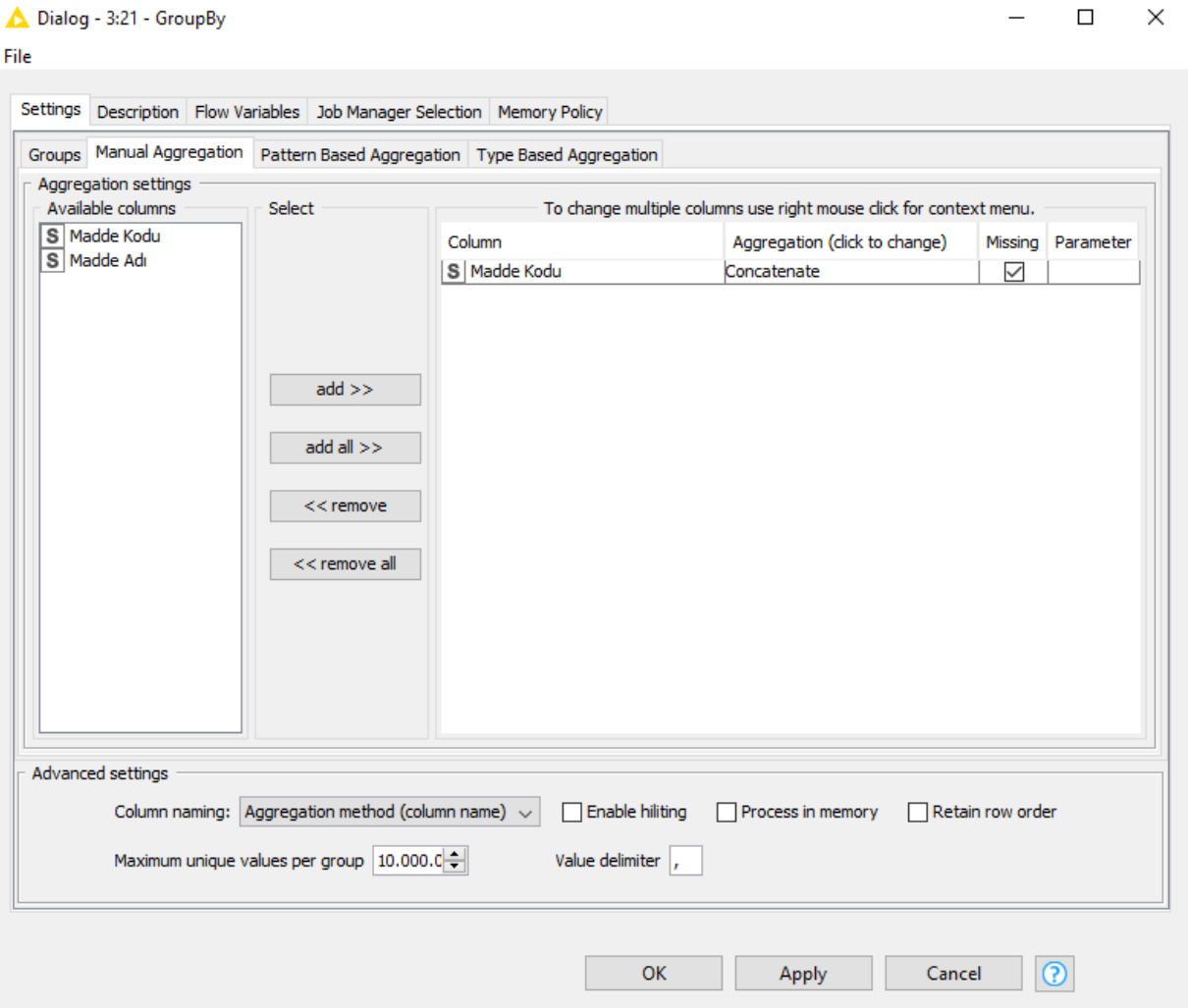
GroupBy Node eklendikten sonra Configure ayarları yapılması gerekmektedir. GroupBy Node eklenmesi ve bağlantısı yapılması sonrası Şekil 5.15.'deki gibi sarı renkli olmaktadır. Yapılan çalışmada Sipariş No kolonuna göre gruplama yapılacaktır. Eğer data Tablo 5.3.'de bulunan madde kodları gibi alt alta ise bunu GroupBy Configure ayarları ile Birliktelik analizine uygun olarak yan yana getirilecektir.



Şekil 29 KNIME GroupBy Node Configure Grups ayarı

GroupBy Node Configure ayar ekranı Şekil 5.16.'daki gibidir. Grups içerisinde gruplamak istenen kolon seçilir. Çalışmamızda Sipariş No'ya göre gruplama yapılacağından Group Column(s) bölümüne Sipariş No kolonu aktarılıp diğer kolonlar alınmamıştır.

Grups penceresinin altında bulunan “Value delimiter” kısmında ise birleştirilecek olan madde kodları arasındaki ayracın virgül olacağını belirtmektedir.



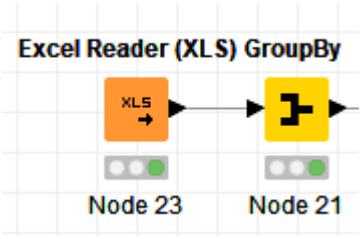
Şekil 30 KNIME GroupBy Node Configure Manual Aggregation ayarı

Aynı sipariş numaralarındaki madde kodlarının yan yana olması için Şekil 5.17.’daki gibi GroupBy Node Configure Manual Aggregation içerisine girilmektedir. Madde Kodu seçilerek Add butonundan Column içerisine alınmaktadır.

Madde Kodu Column altında görülmesi sonrası Şekil 5.17.’deki gibi Missing kutucuğu işaretlenmektedir.

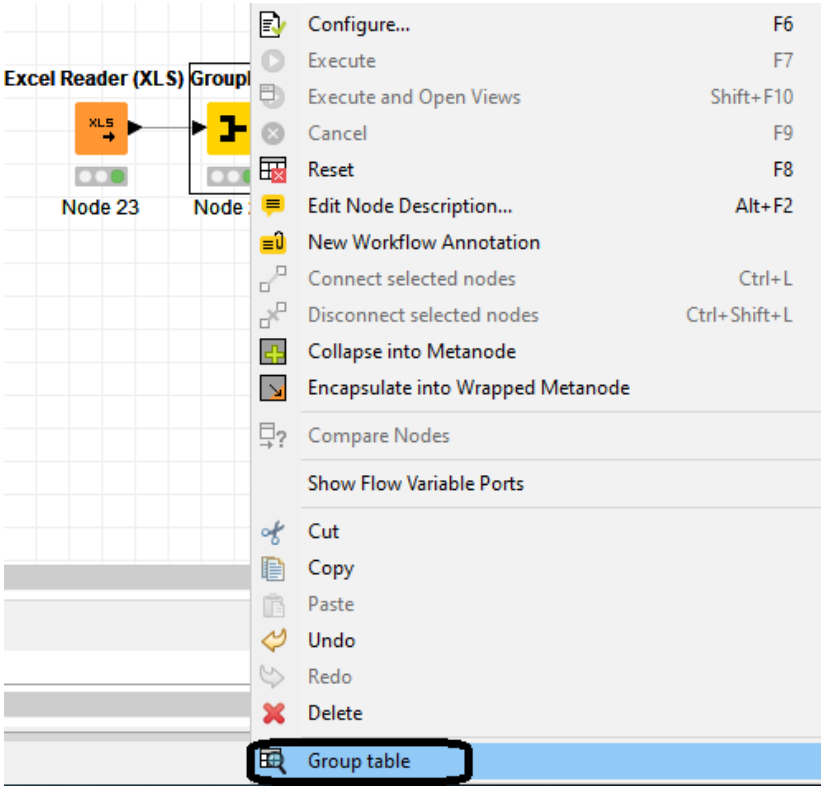
Ayarların tamamlanması sonrası GroupBy üzerinde Mouse ile sağ tuş yaparak gelen seçeneklerden, Execute tıklanarak GroupBy çalıştırılmaktadır.

GroupBy üzerinde Execute yaptıktan sonra Şekil 5.18.'deki gibi yeşil olmaktadır.

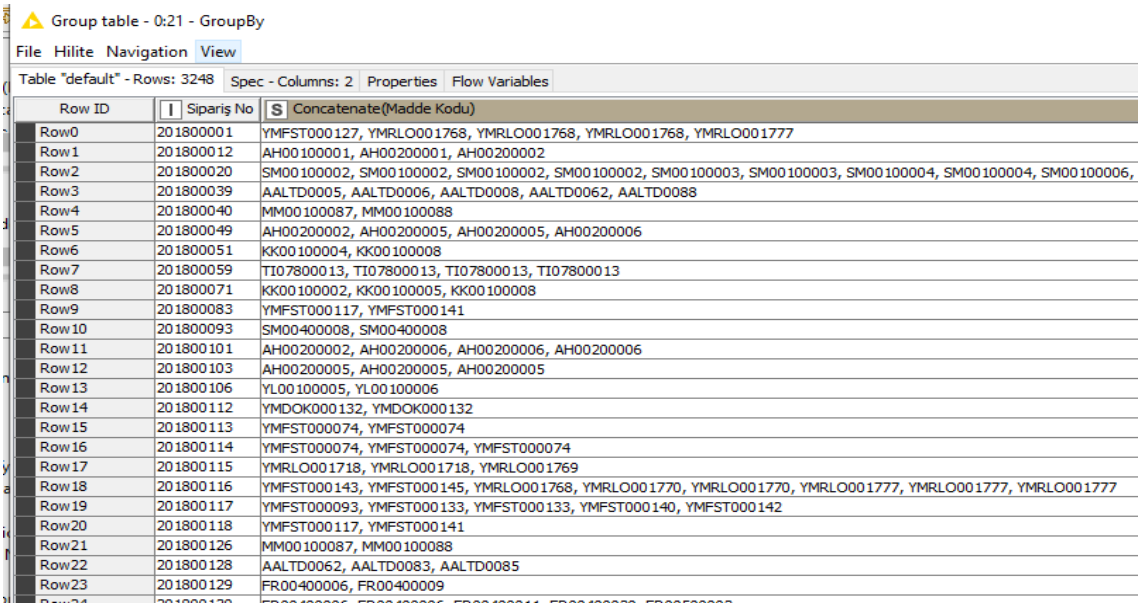


Şekil 31 KNIME GroupBy Node Configure ayarları sonrası

GrupBy yeşil olması sonrası Şekil 5.19.'daki gibi GrupBy üzerinde sağ tuş yaparak, Group table seçeneğinden Şekil 5.20.'deki gibi verinin oluşan son hali gelecektir.



Şekil 32 KNIME GroupBy Node Group table seçim ekranı



Group table - 0:21 - GroupBy

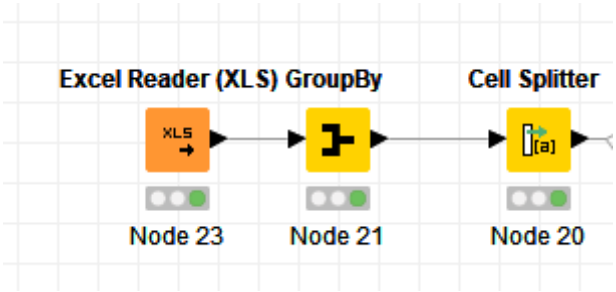
File Hilite Navigation View

Table "default" - Rows: 3248 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	Sipariş No	Concatenate(Madde Kodu)
Row0	201800001	YMFST000127, YMRLO001768, YMRLO001768, YMRLO001768, YMRLO001777
Row1	201800012	AH00100001, AH00200001, AH00200002
Row2	201800020	SM00100002, SM00100002, SM00100002, SM00100002, SM00100003, SM00100003, SM00100004, SM00100004, SM00100006,
Row3	201800039	AALTD0005, AALTD0006, AALTD0008, AALTD0062, AALTD0088
Row4	201800040	MM00100087, MM00100088
Row5	201800049	AH00200002, AH00200005, AH00200005, AH00200006
Row6	201800051	KK00100004, KK00100008
Row7	201800059	TI07800013, TI07800013, TI07800013, TI07800013
Row8	201800071	KK00100002, KK00100005, KK00100008
Row9	201800083	YMFST000117, YMFST000141
Row10	201800093	SM00400008, SM00400008
Row11	201800101	AH00200002, AH00200006, AH00200006, AH00200006
Row12	201800103	AH00200005, AH00200005, AH00200005
Row13	201800106	YL00100005, YL00100006
Row14	201800112	YMDOK000132, YMDOK000132
Row15	201800113	YMFST000074, YMFST000074
Row16	201800114	YMFST000074, YMFST000074, YMFST000074
Row17	201800115	YMRLO001718, YMRLO001718, YMRLO001769
Row18	201800116	YMFST000143, YMFST000145, YMRLO001768, YMRLO001770, YMRLO001770, YMRLO001777, YMRLO001777, YMRLO001777
Row19	201800117	YMFST000093, YMFST000133, YMFST000133, YMFST000140, YMFST000142
Row20	201800118	YMFST000117, YMFST000141
Row21	201800126	MM00100087, MM00100088
Row22	201800128	AALTD0062, AALTD0083, AALTD0085
Row23	201800129	FR00400006, FR00400009

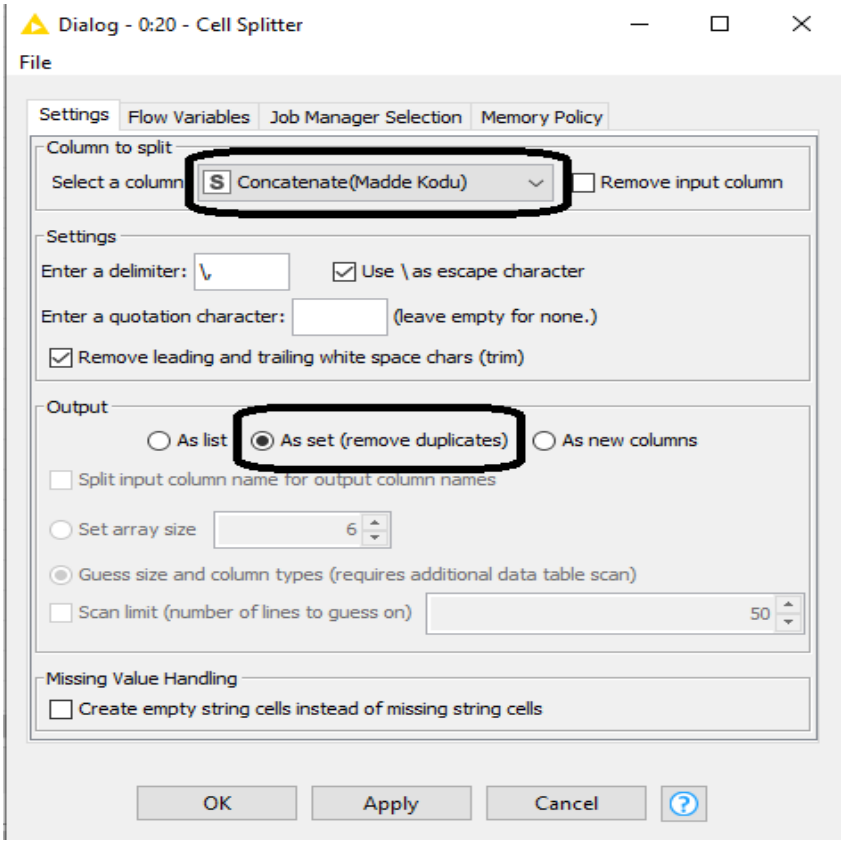
Şekil 33 KNIME GroupBy Node Group Table

Veri Şekil 5.20.'deki gibi olduktan sonra kümeleme işlemi yapılabilmesi için Şekil 5.21.'deki gibi Cell Splitter node'u eklenmelidir.



Şekil 34 KNIME Cell Splitter Node

Cell Splitter node'un Configurasyon ayarları ise Şekil 5.22.'deki gibi yapılmalıdır. Select a column seçeneğinden kümelenmek istenilen kolon seçilir, tekrarlanan veriyi temizlemek için ise as set (remove duplicates) seçeneği seçilmeli ve execute yapılarak çalıştırılmalıdır.



Şekil 35 KNIME Cell Splitter Node Configure

Cell Splitter node çalıştırılması sonrası Şekil 5.23.'deki gibi yeni veri gelmektedir. Yeni gelen bu veri (Concantenate(Madde Kodu)_SplitResultSet) bir sonraki node'a eklenecek olan giriş değeridir. Birliktelik kuralı analizi yapabilmesi için veri (Concantenate(Madde Kodu)_SplitResultSet) kololunundaki formatta olması gerekiyor. Buraya kadar olan adımların yapılma sebebide budur.

Output Table - 0:20 - Cell Splitter

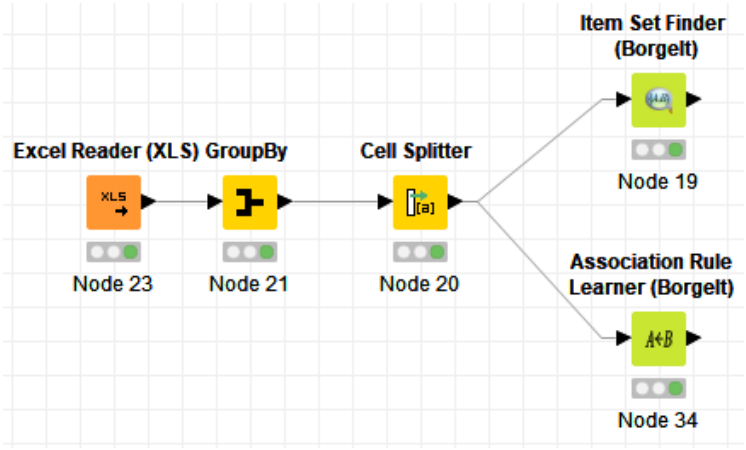
File Hilite Navigation View

Table "default" - Rows: 3248 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	Sipariş No	Concatenate(Madde Kodu)	Concatenate(Madde Kodu)_SplitResultSet
Row0	201800001	YMFST000127, YMRLO001768, YMRLO001768, YMRLO001768, YMRLO001777	[YMFST000127,YMRLO001768,YMRLO001777]
Row1	201800012	AH00100001, AH00200001, AH00200002	[AH00100001,AH00200001,AH00200002]
Row2	201800020	SM00100002, SM00100002, SM00100002, SM00100002, SM00100003, SM0...	[SM00100002,SM00100003,SM00100004,...]
Row3	201800039	AALTD0005, AALTD0006, AALTD0008, AALTD0062, AALTD0088	[AALTD0005,AALTD0006,AALTD0008,...]
Row4	201800040	MM00100087, MM00100088	[MM00100087,MM00100088]
Row5	201800049	AH00200002, AH00200005, AH00200005, AH00200006	[AH00200002,AH00200005,AH00200006]
Row6	201800051	KK00100004, KK00100008	[KK00100004,KK00100008]
Row7	201800059	TI07800013, TI07800013, TI07800013, TI07800013	[TI07800013]
Row8	201800071	KK00100002, KK00100005, KK00100008	[KK00100002,KK00100005,KK00100008]
Row9	201800083	YMFST000117, YMFST000141	[YMFST000117,YMFST000141]
Row10	201800093	SM00400008, SM00400008	[SM00400008]
Row11	201800101	AH00200002, AH00200006, AH00200006, AH00200006	[AH00200002,AH00200006]
Row12	201800103	AH00200005, AH00200005, AH00200005	[AH00200005]
Row13	201800106	YL00100005, YL00100006	[YL00100005,YL00100006]
Row14	201800112	YMDOK000132, YMDOK000132	[YMDOK000132]
Row15	201800113	YMFST000074, YMFST000074	[YMFST000074]
Row16	201800114	YMFST000074, YMFST000074, YMFST000074	[YMFST000074]
Row17	201800115	YMRLO001718, YMRLO001718, YMRLO001769	[YMRLO001718,YMRLO001769]
Row18	201800116	YMFST000143, YMFST000145, YMRLO001768, YMRLO001770, YMRLO00177...	[YMFST000143,YMFST000145,YMRLO001768,...]
Row19	201800117	YMFST000093, YMFST000133, YMFST000133, YMFST000140, YMFST000142	[YMFST000093,YMFST000133,YMFST000140,...]
Row20	201800118	YMFST000117, YMFST000141	[YMFST000117,YMFST000141]
Row21	201800126	MM00100087, MM00100088	[MM00100087,MM00100088]

Şekil 36 KNIME Cell Splitter Output View

Birliktelik kuralında, Destek ve Güven ile Birlikte alınan ürünlerin setleri bulunacağından Şekil 5.24.'deki gibi 2 ayrı node kullanılmıştır.



Şekil 37 KNIME Item Set Finder Node ve Association Rule Learner Node

Destek ve Güven için "Association Rule Learner" node kullanılmıştır. Birlikte alınan ürünlerin bulunabilmesi için ise "Item Set Finder (Borgelt) node'u kullanılmıştır.

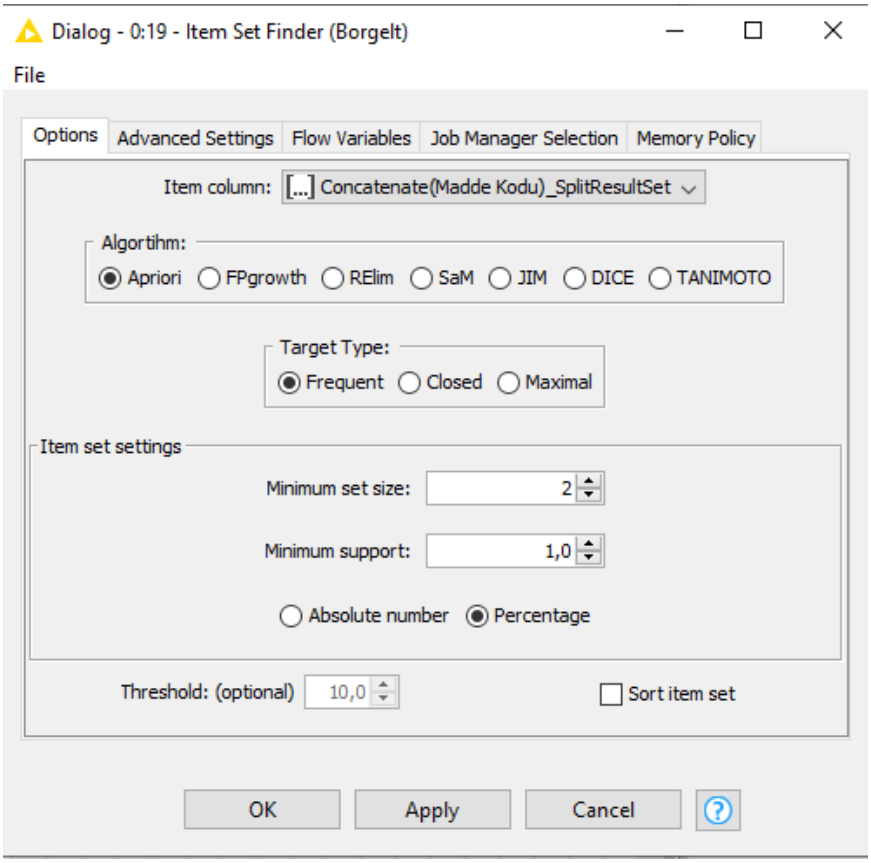
Item Set Finder node'un configurasyon ayarlarında Şekil 5.25.'deki gibi Apriori veya diğer algoritmalardan biri seçilerek birliktelik seti oluşturulabilir.Çalışmada Apriori

algoritması kullanılmıştır.Target type seçeneğinden frequent (ürün seti bazında bir birliktelik) seçilebilir.

Minimum set size alanında, oluşan birlikteliklerden minimum görmek istenilen birliktelik sayısı yazılmalıdır. 1 yazılması durumunda oluşacak listeye tek satınalınan ürünlerde gelmektedir. Çalışmamızda minimum 2 ürünün birlikteliği alınacağından Minimum set size değerine 2 girilmiştir.

Minimum support alanındaysa, en az kaç satınalma ve yukarısı isteniyorsa o rakam girilerek bir sonuç elde edilebilir. Burada dikkat edilmesi gereken, veriye uygun değer girilmesidir. Uygun değer girilmediğinde Item Set Finder Node çalıştırıldığında hata vermektedir. Çalışmada kullanmış olduğumuz veride 1,0 değeri girilmiştir. Yüksek değerlerde hata alınmaktadır.

Absolute number seçeneği, verilerin sayısal olarak listelenmesini sağlamakta iken Percentage ise yüzdelik değerlerini göstermektedir. Çalışmamızda da Percentage seçeneği seçilmiştir.



Şekil 38 Item Set Finder Node Configurasyon Ayarları

Şekil 5.25.'deki ayarlar sonrası execute yapılarak Item Set Finder Node çalıştırıldığında A firması için Şekil 5.26.'daki gibi B firması için ise Şekil 5.27.'deki gibi sonuçlar alınmıştır. Sonuçlara göre 3'lü ve 2'li ürün setleri bulunmuştur.

Item Sets - 0:19 - Item Set Finder (Borgelt)

File Hilite Navigation View

Table "default" - Rows: 4 | Spec - Columns: 4 | Properties | Flow Variables

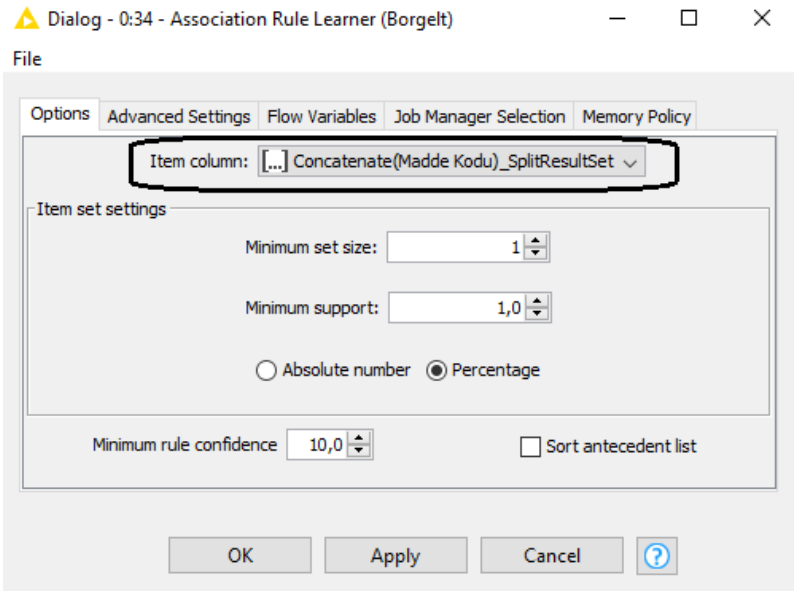
Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row0	[EN00200001,EN00200002,EN00200003]	3	47	1.447
Row1	[EN00200001,EN00200002]	2	47	1.447
Row2	[EN00200001,EN00200003]	2	47	1.447
Row3	[EN00200002,EN00200003]	2	47	1.447

Şekil 39 A Firması Item Set Finder Node Sonuçları

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row0	[503007,503025]	2	11	1.521
Row1	[503007,503008]	2	11	1.521
Row2	[503025,503008]	2	18	2.49
Row3	[OGKV0016,OGSO0026]	2	12	1.66

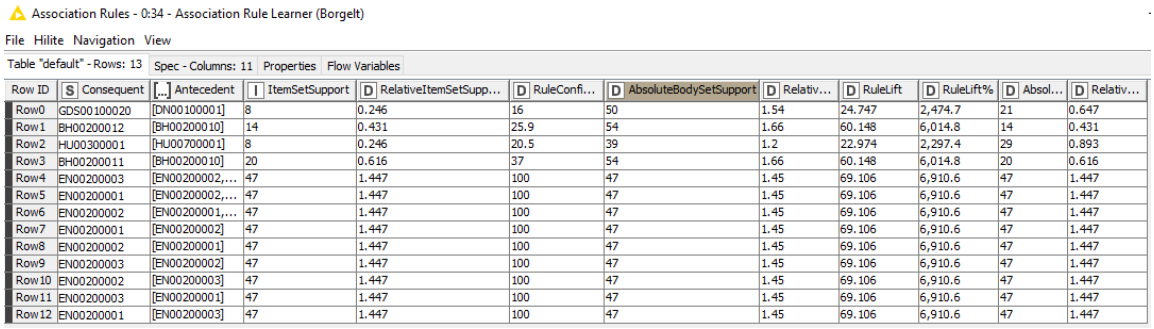
Şekil 40 B Firması Item Set Finder Node Sonuçları

Destek ve Güven için eklenmiş olan Association Rule Learner note ise Apriori algoritmasıyla çalışmaktadır.



Şekil 41.28. Association Rule Learner Note Configure Ekranı

Şekil 6.28.'deki gibi Birliktelik kurallarının bulunabilmesi için Item column alanında madde kodu seçilir. Minimum set size ve minimum support değerleri uygun şekilde girilerek execute ile çalıştırılır. Sonucu görmek için Mouse üzerinden sağ tıklayıp Association Rules seçeneği seçilerek çıkan sonuç Şekil 5.29.'daki gibi görülebilir.



Row ID	S Consequent	[...] Antecedent	I ItemSetSupport	D RelativeItemSetSupp...	D RuleConfi...	D AbsoluteBodySetSupport	D Relativ...	D RuleLift	D RuleLift%	D Absol...	D Relativ...
Row0	GDS00100020	[DN00100001]	8	0.246	16	50	1.54	24.747	2,474.7	21	0.647
Row1	BH00200012	[BH00200010]	14	0.431	25.9	54	1.66	60.148	6,014.8	14	0.431
Row2	HU00300001	[HU00700001]	8	0.246	20.5	39	1.2	22.974	2,297.4	29	0.893
Row3	BH00200011	[BH00200010]	20	0.616	37	54	1.66	60.148	6,014.8	20	0.616
Row4	EN00200003	[EN00200002,...	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row5	EN00200001	[EN00200002,...	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row6	EN00200002	[EN00200001,...	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row7	EN00200001	[EN00200002]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row8	EN00200002	[EN00200001]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row9	EN00200003	[EN00200002]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row10	EN00200002	[EN00200003]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row11	EN00200003	[EN00200001]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447
Row12	EN00200001	[EN00200003]	47	1.447	100	47	1.45	69.106	6,910.6	47	1.447

Şekil 42 A Firması Association Rule Learner Note Sonuçları

Şekil 5.29.'daki veriler ve anlamları; Consequent kısmında bulunan madde kodları, antecedent kısmındaki madde kodları ile birliktelik kuralları karşılaştırılmaktadır. İtem set support kısmında, o madde kodlarının kaç defa birlikte satın alındığını göstermektedir. Diğer bir değişle Consequent alanındaki tekil madde kodlarının Antecedent alanındaki madde kodu setiyle birlikte satın alındığı adedidir. Rule confidence madde kodlarının kendi aralarındaki güven yüzdesini göstermektedir.

Rule Lift (Kaldıraç): Birliktelik kuralları, birliktelikleri belirlerken destek ve güven değerlerini kullanarak sonuç almaktadır. Fakat destek ve güven değerlerinin yüksek olması halinde, kaldıraç değeri ön plana gelmektedir. Rule lift değeri, güven değerinin destek değerine bölünmesiyle oluşan bir katsayı değeridir. Aşağıdaki formül ile ifade edilebilir.

$$\text{Kaldıraç (A} \rightarrow \text{B)} = Güven(A \rightarrow B)/Destek\ (B)$$

Row ID	S Consequent	[...] Antecedent	I ItemSetSupport	D RuleConfidence%	D RuleLift
Row0	GDS00100020	[DN00100001]	8	16	24.747
Row1	BH00200012	[BH00200010]	14	25.9	60.148
Row2	HU00300001	[HU00700001]	8	20.5	22.974
Row3	BH00200011	[BH00200010]	20	37	60.148
Row4	EN00200003	[EN00200002,EN00200001]	47	100	69.106
Row5	EN00200001	[EN00200002,EN00200003]	47	100	69.106
Row6	EN00200002	[EN00200001,EN00200003]	47	100	69.106
Row7	EN00200001	[EN00200002]	47	100	69.106
Row8	EN00200002	[EN00200001]	47	100	69.106
Row9	EN00200003	[EN00200002]	47	100	69.106
Row10	EN00200002	[EN00200003]	47	100	69.106
Row11	EN00200003	[EN00200001]	47	100	69.106
Row12	EN00200001	[EN00200003]	47	100	69.106

Şekil 43 A Firması Association Rule Learner Note Sonuçları - En Fazla Satınalınan Ürüne Göre Destek ve Güven

En fazla satın alınan ürünlere göre birliktelikler Şekil 5.30. üzerinden kontrol edildiğinde "GDS00100020" madde kodu ürününden alanların "DN000100001" ürününden alma olasılığı %16'dır. RuleLift değeri ise 24.747 olmuştur.

Resim 6.30.'da ki listede güven değeri 100 olan birliktelikler incelendiğinde ise "EN00200003" madde kodu ürününden alanların "EN00200002, EN00200001" ürünü alma olasılığı %100 dür.

En fazla satın alınan madde kodlarına bakıldığında ise Şekil 5.31.'deki "EN0020003", "EN0020002" ve "EN0020001" madde kodlarının kendi aralarındaki birliktelikleri olmuştur.

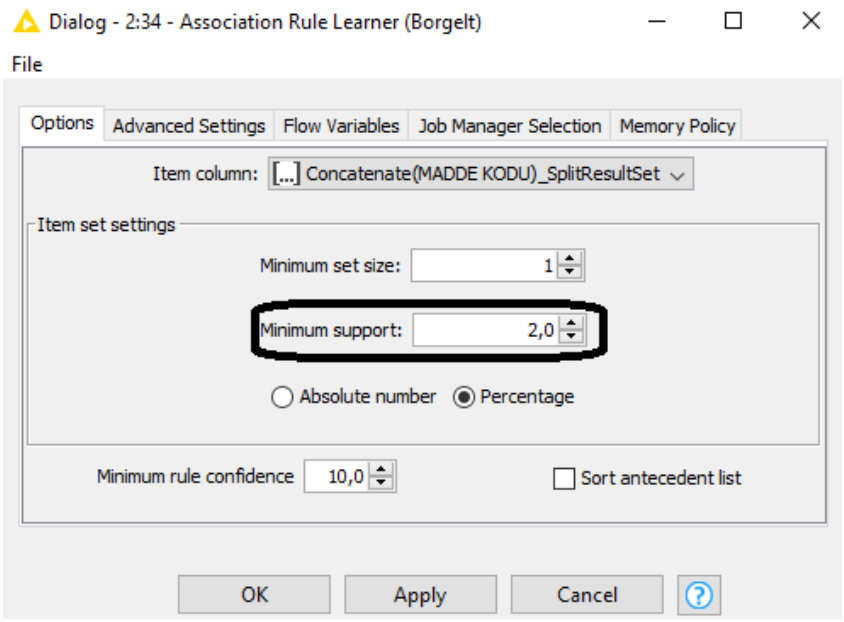
Table "default" - Rows: 13 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	Consequent	Antecedent	ItemSetSupport
Row0	GDS00100020	[DN00100001]	8
Row1	BH00200012	[BH00200010]	14
Row2	HU00300001	[HU00700001]	8
Row3	BH00200011	[BH00200010]	20
Row4	EN00200003	[EN00200002,EN00200001]	47
Row5	EN00200001	[EN00200002,EN00200003]	47
Row6	EN00200002	[EN00200001,EN00200003]	47
Row7	EN00200001	[EN00200002]	47
Row8	EN00200002	[EN00200001]	47
Row9	EN00200003	[EN00200002]	47
Row10	EN00200002	[EN00200003]	47
Row11	EN00200003	[EN00200001]	47
Row12	EN00200001	[EN00200003]	47

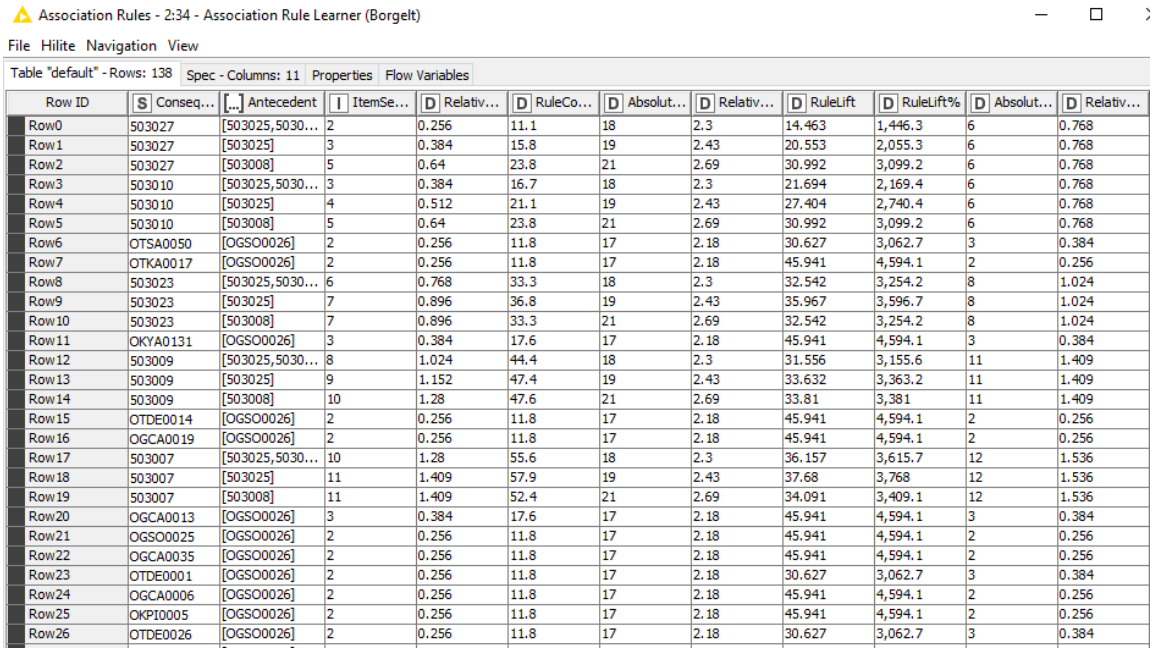
Şekil 44 A Firmasında Birlikte En Fazla Satın Alınan Ürünler

B firması Association Rule Learner Note Sonuçları;

B firması Association Rule Learner Note Configure ayarlarında A firmasına göre minimum support değeri Şekil 5.32.'deki gibi 2,0 olarak girilmiştir. Sonuçlar ise Şekil 5.33.'deki gibi şekillenmiştir.



Şekil 45 B Firması Association Rule Learner Note Configure Ekranı



Association Rules - 2:34 - Association Rule Learner (Borgelt)

File Hilite Navigation View

Table "default" - Rows: 138 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	Conseq...	Antecedent	ItemSe...	Relativ...	RuleCo...	Absolut...	Relativ...	RuleLift	RuleLift%	Absolut...	Relativ...
Row0	503027	[503025,5030...	2	0.256	11.1	18	2.3	14.463	1,446.3	6	0.768
Row1	503027	[503025]	3	0.384	15.8	19	2.43	20.553	2,055.3	6	0.768
Row2	503027	[503008]	5	0.64	23.8	21	2.69	30.992	3,099.2	6	0.768
Row3	503010	[503025,5030...	3	0.384	16.7	18	2.3	21.694	2,169.4	6	0.768
Row4	503010	[503025]	4	0.512	21.1	19	2.43	27.404	2,740.4	6	0.768
Row5	503010	[503008]	5	0.64	23.8	21	2.69	30.992	3,099.2	6	0.768
Row6	OTSA0050	[OGSO0026]	2	0.256	11.8	17	2.18	30.627	3,062.7	3	0.384
Row7	OTKA0017	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row8	503023	[503025,5030...	6	0.768	33.3	18	2.3	32.542	3,254.2	8	1.024
Row9	503023	[503025]	7	0.896	36.8	19	2.43	35.967	3,596.7	8	1.024
Row10	503023	[503008]	7	0.896	33.3	21	2.69	32.542	3,254.2	8	1.024
Row11	OKYA0131	[OGSO0026]	3	0.384	17.6	17	2.18	45.941	4,594.1	3	0.384
Row12	503009	[503025,5030...	8	1.024	44.4	18	2.3	31.556	3,155.6	11	1.409
Row13	503009	[503025]	9	1.152	47.4	19	2.43	33.632	3,363.2	11	1.409
Row14	503009	[503008]	10	1.28	47.6	21	2.69	33.81	3,381	11	1.409
Row15	OTDE0014	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row16	OGCA0019	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row17	503007	[503025,5030...	10	1.28	55.6	18	2.3	36.157	3,615.7	12	1.536
Row18	503007	[503025]	11	1.409	57.9	19	2.43	37.68	3,768	12	1.536
Row19	503007	[503008]	11	1.409	52.4	21	2.69	34.091	3,409.1	12	1.536
Row20	OGCA0013	[OGSO0026]	3	0.384	17.6	17	2.18	45.941	4,594.1	3	0.384
Row21	OGSO0025	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row22	OGCA0035	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row23	OTDE0001	[OGSO0026]	2	0.256	11.8	17	2.18	30.627	3,062.7	3	0.384
Row24	OGCA0006	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row25	OKPI0005	[OGSO0026]	2	0.256	11.8	17	2.18	45.941	4,594.1	2	0.256
Row26	OTDE0026	[OGSO0026]	2	0.256	11.8	17	2.18	30.627	3,062.7	3	0.384

Şekil 46 B Firması Association Rule Learner Note Sonuçları

En fazla satın alınan ürünlere göre birliktelikler Şekil 5.34. üzerinden kontrol edildiğinde "503008" ürününden alınların "503025" ürününden alma olasılığı %94.7'dir. RuleLift değeri ise 35.233 olmuştur. Güven değeri en yüksek "503008" ile "503025" birlikteliklerinin %94.7 olmuştur.

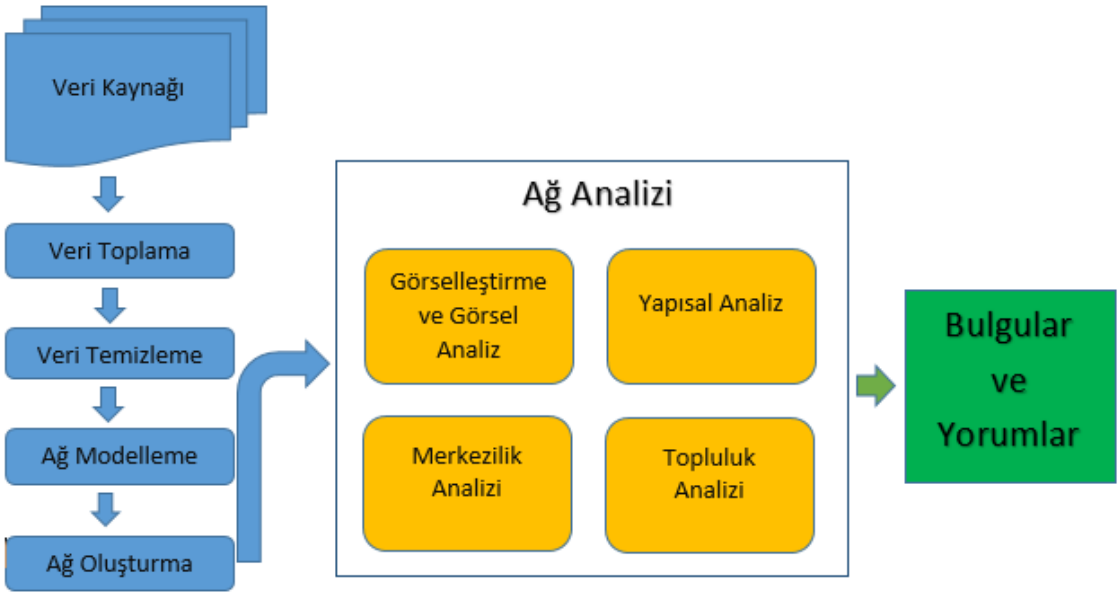
Row ID	Consequent	Antecedent	ItemSetSupport	RuleConfidence%	RuleLift
Row28	503008	[503025]	18	94.7	35.233
Row29	503025	[503008]	18	85.7	35.233
Row137	OGKV0016	[OGSO0026]	12	70.6	36.753
Row18	503007	[503025]	11	57.9	37.68
Row19	503007	[503008]	11	52.4	34.091
Row14	503009	[503008]	10	47.6	33.81
Row17	503007	[503025,503008]	10	55.6	36.157
Row13	503009	[503025]	9	47.4	33.632
Row12	503009	[503025,503008]	8	44.4	31.556
Row133	OGSE0004	[OGSO0026]	8	47.1	40.837
Row135	OTSA0045	[OGSO0026]	8	47.1	33.412
Row136	OGKV0011	[OGSO0026]	8	47.1	33.412
Row9	503023	[503025]	7	36.8	35.967
Row10	503023	[503008]	7	33.3	32.542
Row126	OGCA0031	[OGSO0026]	7	41.2	35.732

Şekil 47 B Firmasında Birlikte En Fazla Satın Alınan Ürünler

BÖLÜM 6. II. YÖNTEM – KARMAŞIK AĞ ANALİZİ

6.1.Karmaşık Ağ Analizi

Karmaşık Ağ Analizi yapmadan önce ham verilerin öncelikle SQL vb. gibi farklı bir program yardımı ile analizi uygun hale getirilmelidir. Karmaşık Ağ Analizinde yapılacak işlemler; veri toplama, veri temizleme, ağın modellenmesi ve oluşturulması süreçleridir. Karmaşık ağ analizi yöntemlerinin akış diyagramı Şekil 6.1.'de verilmiştir.



Şekil 48 Karmaşık Ağ Analizi Yöntemleri

Karmaşık ağ analizinde, ilk veri toplama ile veri temizleme aşaması yapılarak, ağın meydana gelmesi için gerekli veri elde edilir. Ağ modellenmesi sonrası ilgili veri kullanılarak ağ oluşturulur. Ağın oluşması sonrası ise ağ görselleştirilir ve ağ da mevcut olan yapısal desenler aranarak görsel analiz yapılır. Merkezilik ve topluluk analizi sonrası ağın karakteristik özellikleri hakkında bilgi almamıza yarayan yapısal analiz yapılır. En son işlem olarak ise elde edilen bulgular ile karmaşık ağ yorumlanır.

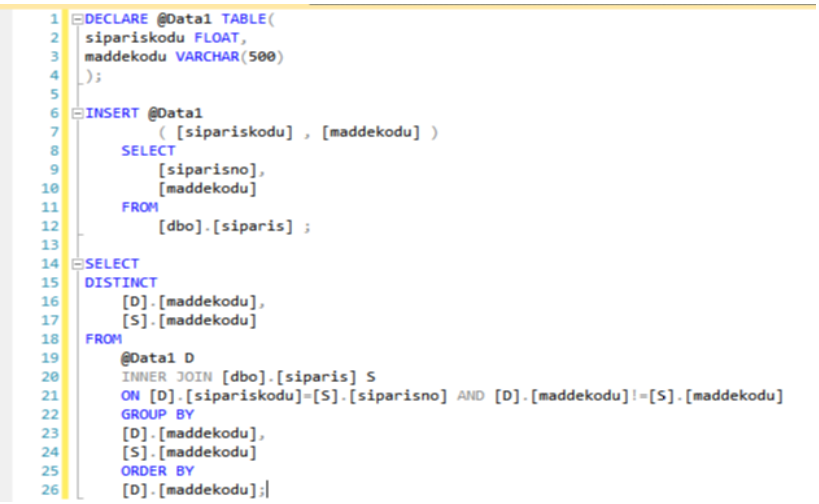
6.2.Verilerin Karmaşık Ağ Analizine Uygun Hale Getirilmesi

Çalışmada kullanılan 2 firmanın ham verileri Tablo 6.1.'deki gibidir.

Tablo 6.1. A ve B Firması Ham Verileri

A Firması Verileri		B Firması Verileri	
Sipariş No	**Madde Kodu**	**Sipariş No**	**Madde Kodu**
201800001	YMFST000127	SA00113921	431313
201800001	YMRLO001768	SA00113921	432161
201800001	YMRLO001768	SA00113921	432169
201800001	YMRLO001768	SA00113921	433024
201800001	YMRLO001777	SA00113930	353008
201800012	AH00100001	SA00113930	353019
201800012	AH00200001	SA00113938	493128
201800012	AH00200002	SA00113938	493130
201800020	SM00100002	SA00113956	502068
201800020	SM00100002	SA00113956	502068
201800020	SM00100002	SA00113956	502069
201800020	SM00100002	SA00113956	502070
201800020	SM00100003	SA00113962	OTDE0032
201800020	SM00100003	SA00113962	OTSA0043
201800020	SM00100004	SA00113962	OTSA0064
201800020	SM00100004	SA00113974	503035
201800020	SM00100006	SA00113974	503036
201800020	SM00100006	SA00113974	503038
201800039	AALTD0005	SA00113974	503039
201800039	AALTD0006	SA00113974	503044
201800039	AALTD0008	SA00113974	503045
201800039	AALTD0062	SA00113974	503152
201800039	AALTD0088	SA00113974	503180
201800040	MM00100087	SA00113975	503052
201800040	MM00100088	SA00113975	503060

SQL programı yardımı ile çalışma verileri One mode Projection yöntemi kullanılarak sosyal ağ analizine uygun hale getirilmiştir. SQL ekran görüntüsü Şekil 6.2.'deki gibidir.



```
DECLARE @Data1 TABLE(
sipariskodu FLOAT,
maddekodu VARCHAR(500)
);

INSERT @Data1
        ( [sipariskodu] , [maddekodu] )
    SELECT
        [siparisno],
        [maddekodu]
    FROM
        [dbo].[siparis] ;

SELECT
DISTINCT
    [D].[maddekodu],
    [S].[maddekodu]
FROM
    @Data1 D
    INNER JOIN [dbo].[siparis] S
    ON [D].[sipariskodu]=[S].[siparisno] AND [D].[maddekodu]!=[S].[maddekodu]
    GROUP BY
    [D].[maddekodu],
    [S].[maddekodu]
    ORDER BY
    [D].[maddekodu];
```

Şekil 49 SQL üzerinde verilerin sosyal ağ analizine uygun hale getirilmesi

Veriler SQL çalışması sonrası sosyal ağ analizine uygun hale getirilmesiyle A firmasında Tablo 6.2.'deki gibi, B firmasında ise Tablo 6.3.'deki gibi eşlemeler meydana gelmiştir. Birebir eşlemeler, aynı sipariş numarası içerisinde bulunan madde kodlarından oluşmuştur. Örnek; AALTD0006 madde kodu, AALTD0008 madde kodu ile 201800039 sipariş numarası içerisinde birbikte satınalınmış olması sebebiyle eşleme olmuştur. Diğer tüm eşlemelerin aynı sipariş numara içesinde geçtiği kontrol edilmiş olup eşlemerin doğru olduğu görülmektedir.

Tablo 6.2. A firması verisi ve SQL sonuçları

A Firması Verisi	
Sipariş No	**Madde Kodu**
201800001	YMFST000127
201800001	YMRLO001768
201800001	YMRLO001768
201800001	YMRLO001768
201800001	YMRLO001777
201800012	AH00100001
201800012	AH00200001
201800012	AH00200002
201800020	SM00100002
201800020	SM00100002
201800020	SM00100002
201800020	SM00100002
201800020	SM00100003
201800020	SM00100003
201800020	SM00100004
201800020	SM00100004
201800020	SM00100006
201800020	SM00100006
201800039	AALTD0005
201800039	**AALTD0006**
201800039	**AALTD0008**
201800039	AALTD0062
201800039	AALTD0088
201800040	MM00100087
201800040	MM00100088

A Firması SQL Sonuçları	
Madde Kodu 1	**Madde Kodu 2**
AA00300001	AA00300002
AALTD0001	AALTD0003
AALTD0001	AALTD0007
AALTD0006	AALTD0007
AALTD0001	AALTD0008
AALTD0006	**AALTD0008**
AALTD0002	AALTD0009
AALTD0005	AALTD0009
AALTD0007	AALTD0009
AALTD0003	AALTD0035
AALTD0008	AALTD0035
AALTD0002	AALTD0037
AALTD0009	AALTD0037
AALTD0007	AALTD0042
AALTD0001	AALTD0044
AALTD0006	AALTD0044
AALTD0042	AALTD0044
AALTD0002	AALTD0046
AALTD0005	AALTD0046
AALTD0009	AALTD0046
AALTD0003	AALTD0048
AALTD0008	AALTD0048
AALTD0035	AALTD0048
AALTD0037	AALTD0048
AALTD0044	AALTD0048

Tablo 6.3. B firması verisi ve SQL sonuçları

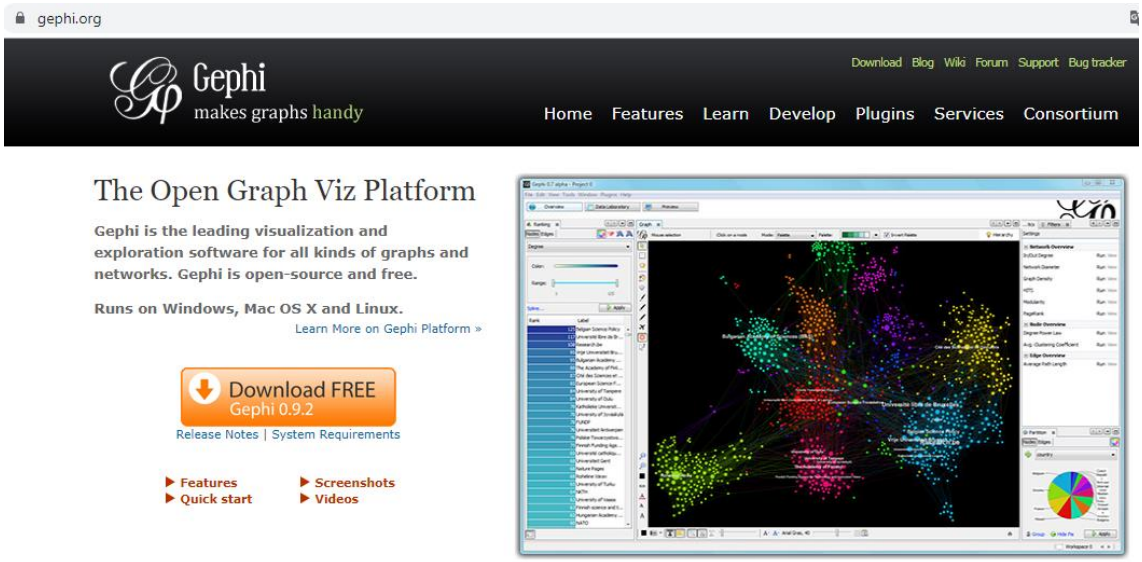
B Firması Verisi		B Firması SQL Sonuçları	
Sipariş No	**Madde Kodu**	**Madde Kodu 1**	**Madde Kodu 2**
SA00113921	**431313**	432092	433019
SA00113921	432161	432194	433019
SA00113921	432169	**431313**	**433024**
SA00113921	**433024**	432161	433024
SA00113930	353008	432169	433024
SA00113930	353019	433046	433047
SA00113938	493128	433046	433048
SA00113938	493130	433047	433048
SA00113962	OTDE0032	433046	433049
SA00113962	OTSA0043	433047	433049
SA00113962	OTSA0064	433048	433049
SA00113974	503035	433046	433050
SA00113974	503036	433047	433050
SA00113974	503038	433048	433050
SA00113974	503039	433049	433050
SA00113974	503044	433046	433051
SA00113974	503045	433047	433051
SA00113974	503152	433048	433051
SA00113974	503180	433049	433051
SA00113975	503052	433050	433051
SA00113975	503060	433004	433053
SA00113979	441130	433017	433053
SA00113979	441154	433004	433054
SA00113979	441183	433017	433054
SA00113979	441188	433053	433054

Ham veri SQL yardımı ile ağ analizine uygun hale getirilmesi sonrası gerekli çalışma Gephi programı ile yapılmıştır.

6.3.GEPHI Uygulaması ve Detayları

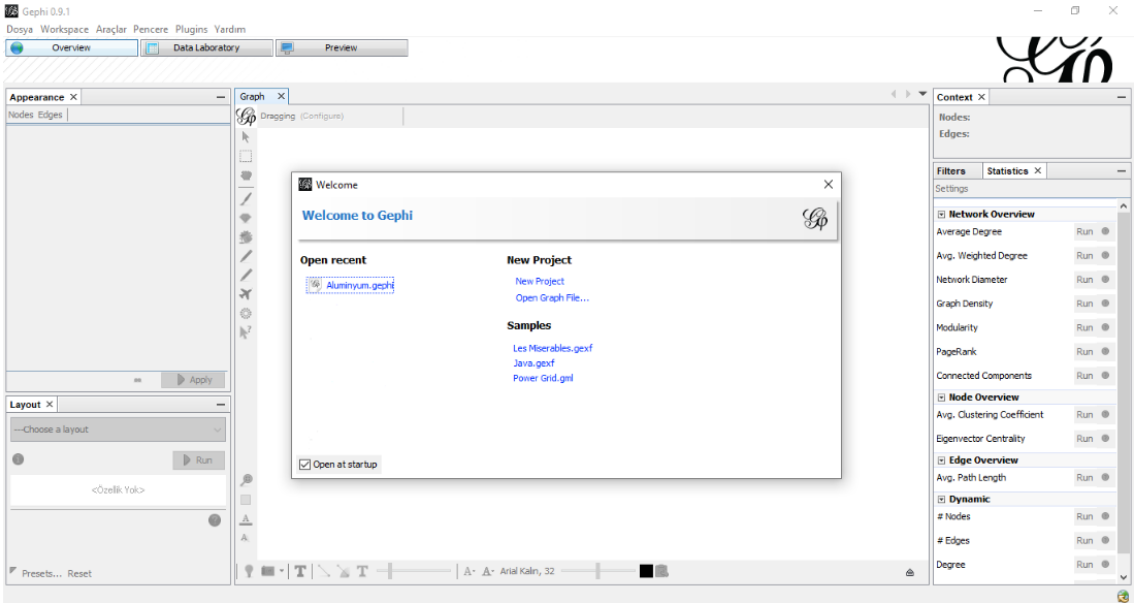
Gephi, açık kaynaklı bir yazılım olup ücretsiz bir programdır. Ağ görselleştirme ve analizi yapmakta ve java tabanlı çalışmaktadır. Tüm işletim sistemlerinde çalışmaktadır [50]. İlk olarak Fransa'da Compiègne Teknoloji Üniversitesi (UTC) öğrencileri tarafından geliştirilmiştir. İlk Sürümü 2008 yılında yayınlanmış ve 2010 yılında Gephi'in gelişimi için Fransa'da kar amacı gütmeyen Gephi Konsorsiyumu kurulmuştur. Gephi form sitesi üzerinde çok sayıda blog ve makale üreten geniş bir kullanıcı kitlesi tarafından desteklenmektedir [51].

Gephi uygulamasının adresi https://gephi.org şeklindedir. Sitenin giriş sayfası Şekil 6.3.'deki gibi olup Download Free seçeneğinden program indirilebilir.



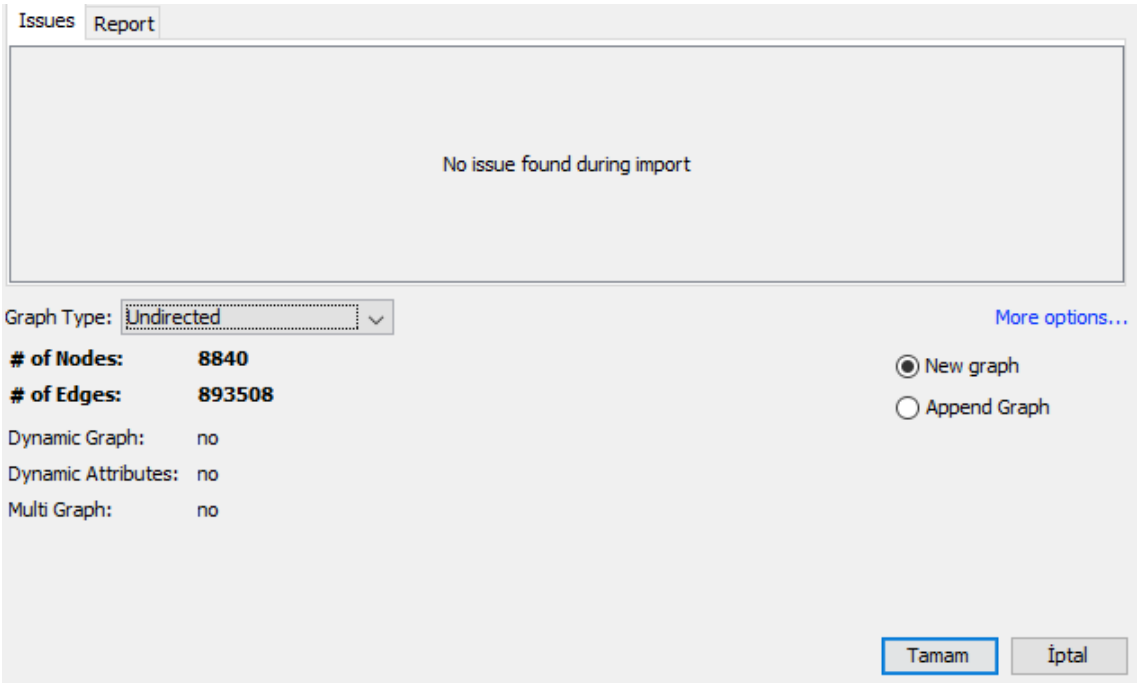
Şekil 50 Gephi adresinin giriş ekranı

Gephi kurulumu sonrası programın açılış ekranı Şekil 6.4.'deki gibidir. Open recent bölümünde daha önce yapılan çalışmalar gelmektedir. Yeni bir çalışma yaratmak için New Project seçeneği kullanılmaktadır. Samples kısmında ise Gephi ile beraber gelen örnek çalışmalar bulunmaktadır.



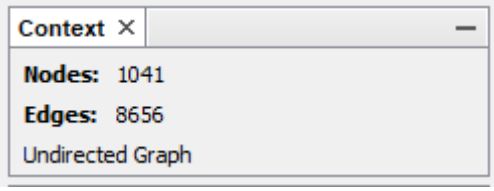
Şekil 51 Gephi giriş ekranı

New Project seçeneğinden A firmasının verileri Şekil 6.5.'deki gibi yönsüz (Undirected) yüklendiğinde 8840 düğüm ve 893508 kenar oluşmuştur.



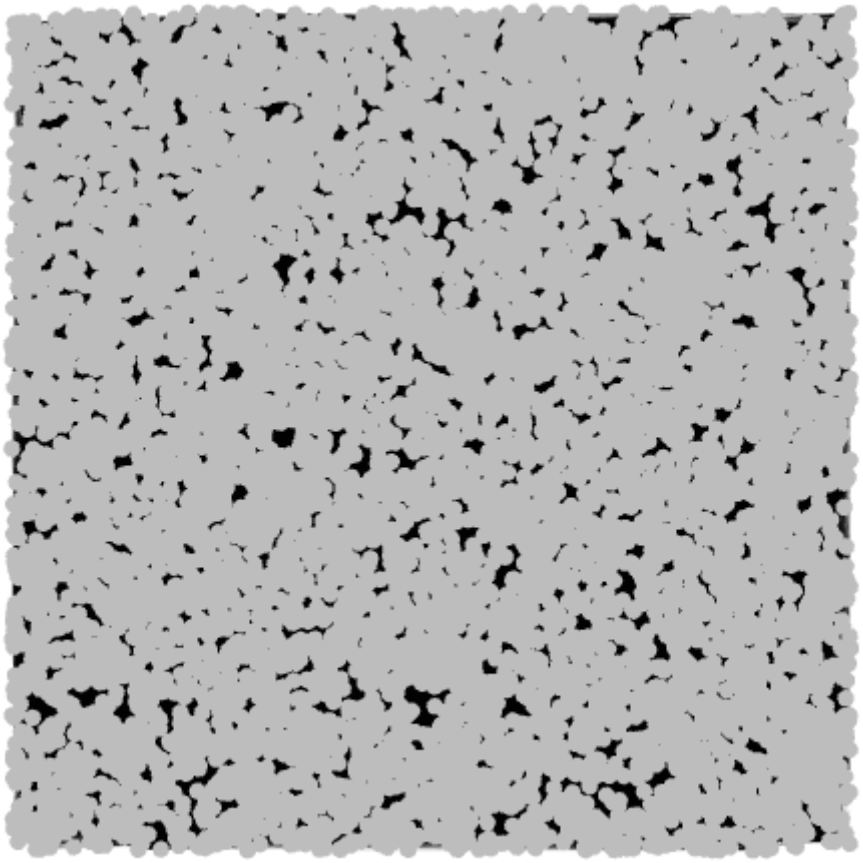
Şekil 52 A firması verilerinin Gephi programına eklenmesi

B firmasının verileri ise Şekil 6.6.'daki gibi yönsüz (Undirected) yüklendiğinde 1041 düğüm ve 8656 kenar oluşmuştur.



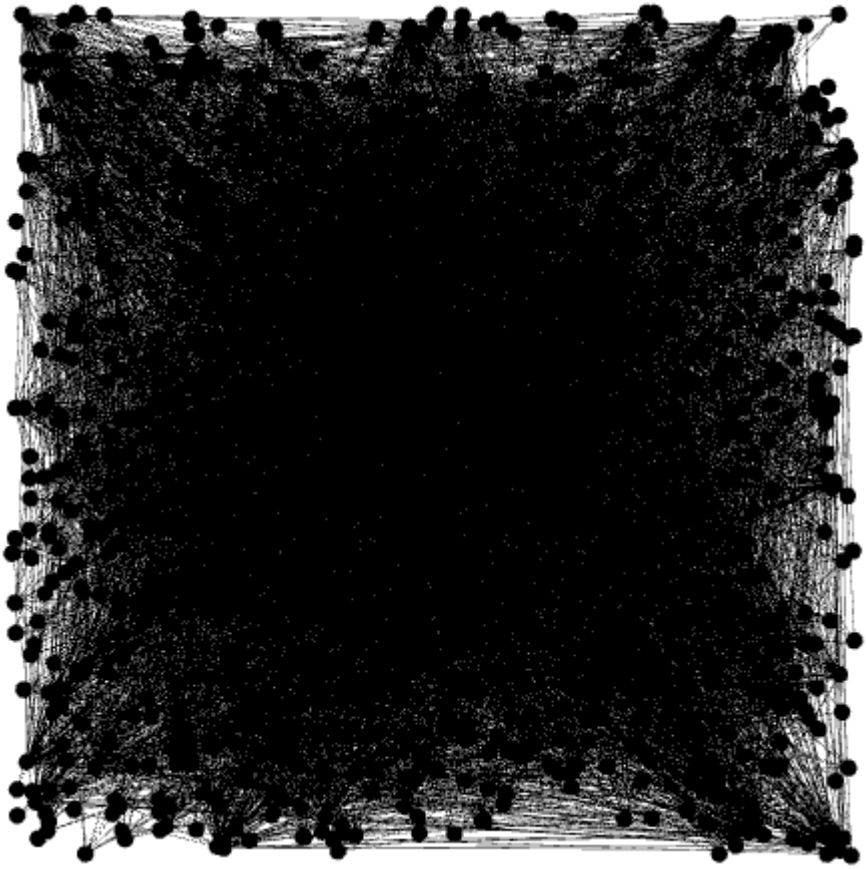
Şekil 53 B firması düğüm ve kenar sayısı

A firmasının verileri yüklemesi sonrası rasgele oluşan karmaşık ağ Şekil 6.7.'deki gibidir. Düğüm ve kenar sayılarının çok yoğun olması sebebiyle iç içe karmaşık bir ağ meydana gelmiştir.



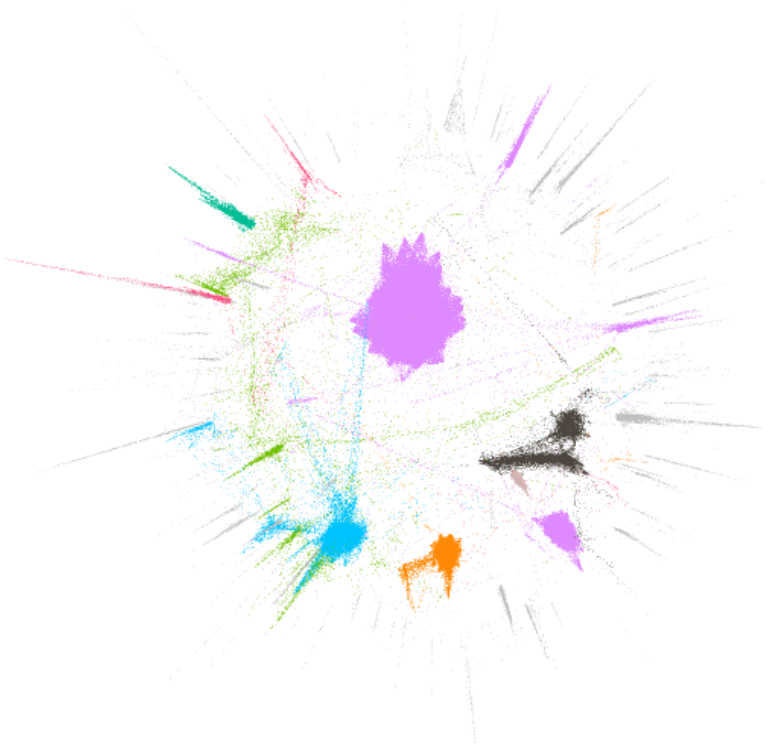
Şekil 54 A firması verisinin rasgele ağ yerleşimi

B firmasının verileri yüklemesi sonrası rasgele oluşan karmaşık ağ ise Şekil 6.8.'deki gibidir.

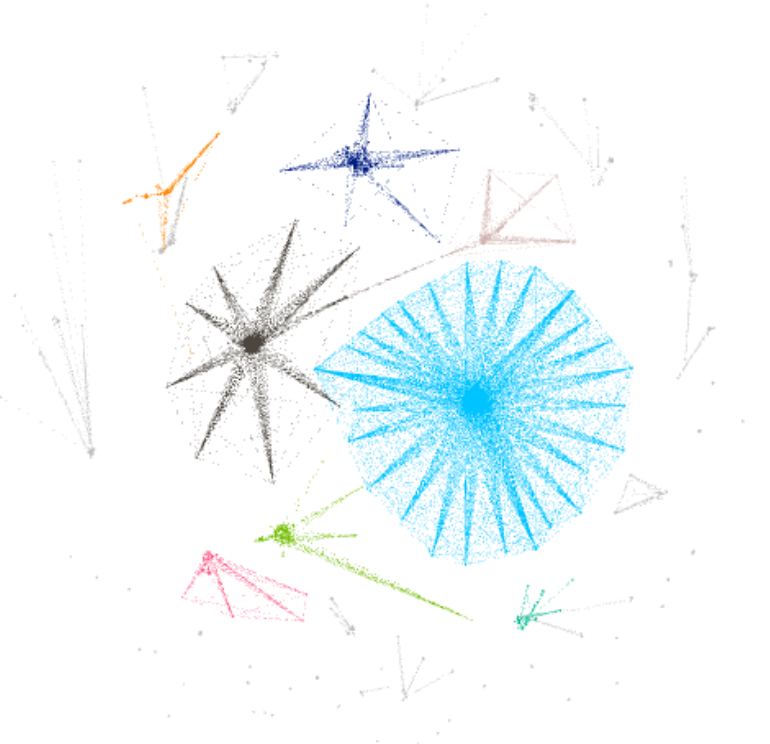


Şekil 55 B firması verisinin rasgele ağ yerleşimi

Yerleşim argoritması ve bölgelerin renklendirilmesi sonrası A firmasında oluşan Şekil 6.9.'daki gibi iken B firmasında oluşan ise Şekil 6.10.'daki gibi olmuştur. Ağ analizi satın alma verileri üzerinden yapılmış olması sebebiyle oluşan ağlar çok parçalı ve karmaşık olarak meydana gelmiştir.



Şekil 56 A Firması Karmaşık Ağın ForceAtlas 2 Görünümü



Şekil 57 B Firması Karmaşık Ağın ForceAtlas 2 Görünümü

6.3.1. Merkezilik Analizi

Ağdaki düğümlerin önemini bulmak için merkezilik analizi uygulanmıştır. Ağdaki bir düğüme ait merkezilik değeri ne kadar yüksek ise düğümün önemi o kadar yüksektir.

Bu analizde iki firma için derece, arasındalık, yakınlık ve özvektör merkezilik ölçütleri yapılmıştır. Analiz sonuçlarına göre ağ da en yüksek değerlere sahip ilk 20 düğümün merkezilik ölçütleri verilmiştir. Değerler sırasıyla Tablo 6.4., Tablo 6.5., Tablo 6.6., Tablo 6.7., Tablo 6.8., Tablo 6.9., Tablo 6.10. ve Tablo 6.11.'de verilmiştir.

Tablo 6.4. A Firmasının Derece Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Derece Merkezilik
1	RY00100272	1.209
2	KO01100144	1.176
3	RY00100382	1.175
4	RY00100555	1.175
5	RY00700030	1.175
6	RY00100041	1.171
7	RY00100315	1.171
8	**RY00100445**	1.171
9	RY00100570	1.171
10	RY00700039	1.171
11	RY01200021	1.171
12	RY00100128	1.170
13	RY00100148	1.170
14	RY00100269	1.170
15	RY00100316	1.170
16	RY00100379	1.170
17	RY00100837	1.170
18	RY00100306	1.169
19	RY00900028	1.169
20	GU00100097	1.168

Tablo 6.5. B Firmasının Derece Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Derece Merkezilik
1	606135	86
2	606136	86
3	606137	86
4	606138	86
5	606091	84
6	606092	84
7	606093	84
8	606094	84
9	606095	84
10	606096	84
11	606004	82
12	606006	82
13	606015	82
14	606016	82
15	606017	82
16	606018	82
17	606019	82
18	606020	82
19	606021	82
20	606022	82

Tablo 6.6. A Firmasının Arasındalık Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Arasındalık Merkezilik
1	GU00300040	622.026,5
2	RY00101478	520.466,824
3	BE01600003	218.690,166
4	BE01600001	218.008
5	TS02000032	206.105,601
6	ID02000002	205.230,085
7	GU01700003	159.323,440
8	HR00900022	129.125,769
9	HR00600002	114.882,786
10	TS12300055	105.688,533
11	TS12300054	103.852,155
12	RY00100291	97.592,494
13	RY00700039	94.384,050
14	**RY00100445**	94.384,050
15	RY00100343	93.698,645
16	RY00100255	93.698,645
17	DC00100010	91.703,306
18	TS02000050	90.866,098
19	RY00100272	85.102,429
20	ID01800013	83.476,098

Tablo 6.7. B Firmasının Arasındalık Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Arasındalık Merkezilik
1	641100	3.330,0
2	641014	1.566,262
3	641133	1.025,606
4	641149	840,458
5	641132	625,083
6	641134	625,083
7	431354	623,693
8	431336	587,668
9	640017	579,073
10	640019	579,073
11	641136	552,419
12	641137	552,419
13	641041	506,034
14	442096	498,395
15	603185	492,600
16	641135	492,0
17	431363	462,757
18	441238	400,881
19	504035	390,0
20	441008	380,724

Tablo 4.8. A Firmasının Yakınlık Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Yakınlık Merkeziliği
1	AA00300001	0,857
2	AA00300002	0,750
3	AALTD0001	0,569
4	AALTD0003	0,617
5	AALTD0007	0,617
6	AALTD0006	0,630
7	AALTD0008	0,630
8	AALTD0002	0,617
9	AALTD0009	0,617
10	AALTD0005	0,630
11	AALTD0035	0,617
12	AALTD0037	0,617
13	AALTD0042	0,707
14	AALTD0044	0,707
15	AALTD0046	0,707
16	AALTD0048	0,617
17	AALTD0062	0,690
18	AALTD0063	0,604
19	AALTD0081	0,358
20	AALTD0082	0,387

Tablo 6.9 B Firmasının Yakınlık Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Yakınlık Merkeziliği
1	421004	1
2	610056	1
3	641053	1
4	641059	1
5	493147	1
6	660120	1
7	419064	1
8	419070	1
9	605050	1
10	621034	1
11	505041	1
12	608109	1
13	493130	1
14	611039	1
15	504143	1
16	660109	1
17	660110	1
18	550016	1
19	419044	1
20	419077	1

Tablo 6.10. A Firmasının Özvektör Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Özvektör
1	RY00100272	1
2	RY00100382	0,999
3	RY00100555	0,999
4	RY00700030	0,999
5	RY00100379	0,999
6	RY01200021	0,999
7	**RY00100445**	0,999
8	RY00700039	0,999
9	RY00100269	0,999
10	RY00100148	0,999
11	GU00100097	0,999
12	RY00101394	0,999
13	RY00100374	0,999
14	RY00100010	0,999
15	RY00100036	0,999
16	RY00100395	0,999
17	RY01500057	0,999
18	RY00100869	0,999
19	RY00900122	0,999
20	RY00100027	0,999

Tablo 6.11. B Firmasının Özvektör Merkeziliğine Göre İlk 20 Madde Kodu

Sıralama	Madde Kodu	Özvektör
1	606135	1
2	606136	1
3	606137	1
4	606138	1
5	606091	0,999
6	606092	0,999
7	606093	0,999
8	606094	0,999
9	606095	0,999
10	606096	0,999
11	606004	0,997
12	606006	0,997
13	606015	0,997
14	606016	0,997
15	606017	0,997
16	606018	0,997
17	606019	0,997
18	606020	0,997
19	606021	0,997
20	606022	0,997

Derece merkezilik sonuçlarına bakıldığında, A firmasının derece merkeziliği tablosunda bulunan ilk madde kodu RY00100272 'in en yüksek dereceye sahip olduğu, diğerlerinin değerlerinin ise birbirlerine çok yakın olduğu gözlemlenmiştir. B firmasında ise ilk dört madde kodunun (606135, 606136, 606137, 606138) derecesinin yüksek olduğu, diğerinin dereceleri ise giderek azaldığı gözlemlenmiştir.

Özvektör merkezilik sonuçlarına bakıldığında, A firması derece merkezinde en yüksek değere sahip olan RY00100272 madde kodunun özvektör merkeziliğinde 1 tam değerle en yüksek değere sahip madde kodu olduğu gözlemlenmiştir. B firmasında ise yine derece merkeziğinde en yüksek değere sahip olan (606135, 606136, 606137, 606138) 4 madde kodunun, 1 tam değerle en yüksek değere sahip olan madde kodları olduğu gözlemlenmiştir.

Arasındalık merkeziliği tablolarına bakıldığında ise A firmasında RY00101478 ve GU00300040 madde kodlarının değerlerinin, diğerlerine göre çok üstün olduğu ve bu madde kodlarının, derece merkeziliğinde ve özvektör merkeziliğinde ilk 20'de olmadığı görülmektedir. B firmasında ise 641100 ve 641014 madde kodlarının değerlerinin

diğerlerine göre çok üstün olduğu ve bu madde kodlarının derece merkeziliğinde ve özvektör merkeziliğinde ilk 20'de olmadığı görülmektedir.

Yakınlık merkeziliği sonuçlarına bakıldığında ise A firmasının ilk sırada bulunan madde kodunun 1'e yakın olduğu, diğer madde kodlarında ise yakınlık merkezi değerinin giderek azaldığı görülmektedir. B firması tablosu incelendiğinde ise ilk 20'de bulunan madde kodlarının tamamının değeri 1 olduğu dikkat çekmektedir.

Birliktelik analizde destek ve güven değerlerine göre en fazla satın alınan madde kodlarının, merkezilik analizinde değerleri en yüksek olan 20 madde kodu arasında bulunmadığı gözlemlenmiştir. Destek ve güven değerlerinde ilk sırada bulunan bulunan madde kodlarının almış olduğu değerlerin kontrol edilmesi sonrası, A firmasının bazı madde kodları değerleri Tablo 6.12.'deki gibidir. B firması madde kodları değerleri ise Tablo 6.13.'deki gibidir.

Tablo 5 A firmasının birliktelik analizindeki madde kodlarının merkezilik analizi

MERKEZİLİK ANALİZİ	**A FİRMASI MADDE KODLARI**				
	HU00300001	**HU00700001**	**EN00200001**	**EN00200002**	**EN00200003**
Derece M.	**4**	**4**	**2**	**2**	**2**
Derece M. Sırası	7828	7829	8333	8530	8334
Arasındalık M.	**0,33**	**0,33**	**0**	**0**	**0**
Arasındalık M. Sırası	2595	2596	4230	8763	4231
Yakınlık M.	**1**	**1**	**1**	**1**	**1**
Yakınlık M. Sırası	548	549	159	1361	160
Özvektör M.	**0,0004**	**0,83**	**0,00002**	**0,00002**	**0,00002**
Özvektör M. Sırası	8002	8003	2231	8472	8473

Tablo 6 B firmasının birliktelik analizindeki madde kodlarının merkezilik analizi

MERKEZİLİK ANALİZİ	B FİRMASI MADDE KODLARI		
	503007	503008	503025
Derece Merkeziliği	**7**	**6**	**7**
Derece M. Sırası	563	616	560
Arasındalık Merkeziliği	**0,166**	**0**	**0,166**
Arasındalık M. Sırası	359	558	363
Yakınlık Merkeziliği	**1**	**0,875**	**1**
Yakınlık M. Sırası	58	59	64
Özvektör Merkeziliği	**0,001297**	**0,001127**	**0,001297**
Özvektör M. Sırası	598	626	602

A firmasının Tablo 6.12.'deki tablo değerleri kontrol edildiğinde en çok dikkat çeken EN00200001, EN00200002 ve EN00200003 isimli madde kodlarıdır. Bu madde kodlarından biri alındığında, diğerinin alınma olasılığı %100'dür. Merkezilik analizinde üçünün değerlerinin aynı olması ve arasındalık merkeziliği değerinin 0 olması, 3 madde kodunun başka ürünlerle alınmadığı ve her alımda hep birlikte alındığını göstermektedir. Bu da ürünlerin sadece kendi içinde küme oluşturduğunu göstermektedir.

B firmasının Tablo 6.13.'deki tablo değerleri kontrol edildiğinde en çok dikkat çeken 503008 isimli madde kodu olmuştur. A firmasındaki gibi 503008 madde kodu, sürekli alım yapılan küçük bir ağ içerisinde bulunmaktadır.

Verilerin büyük bir market verileri olmaması ve farklı sektörlerden veriler olması sebebiyle, market veri sonuçlarındaki gibi yüksek değerler çıkmaması normaldir.

6.3.2. Topluluk Analizi

Satın alma verilerinin topluluk analizi için, Gephi içerisinden Modularity (modülarite) methodu uygulanmıştır. Modülarite methodu ile A firmasında, 6 adet büyük küme, 5 adet küçük küme bulunmaktadır. Bu kümeler Şekil 6.11.'de görülmektedir. B firmasında ise 4 adet büyük küme, 7 adet küçük küme bulunmaktadır. Bu kümeler Şekil 6.12.'de görülmektedir.

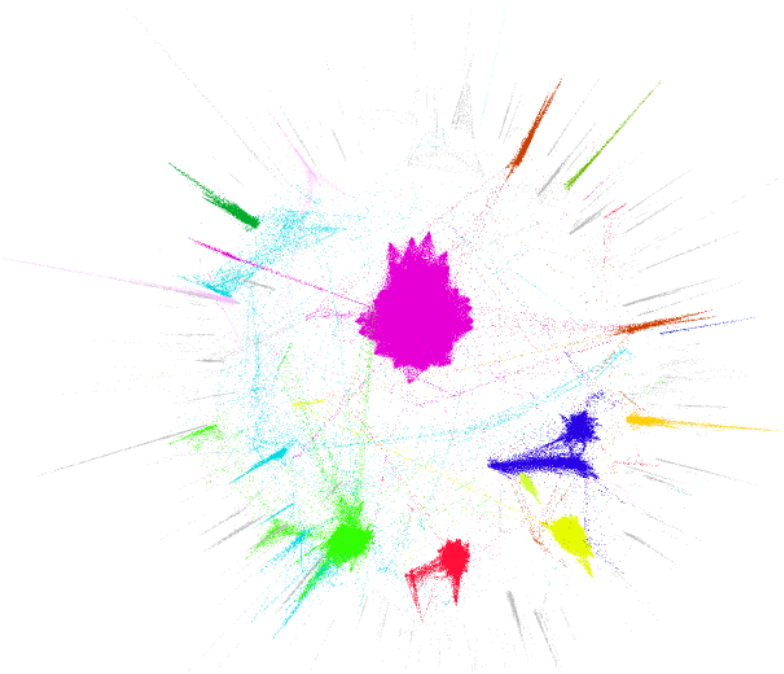
A firmasının ağ da bulunan kümelerinin özellikleri Tablo 6.14.'de, B firmasının ise Tablo 6.15.'de verilmiştir. Tablolarda bulunan yüzdelik değerler her kümenin kendi içerisinde yüzdelik değerleridir.

Tablo 7 A Firması Karmaşık Ağında Bulunan Kümeler ve Özellikleri

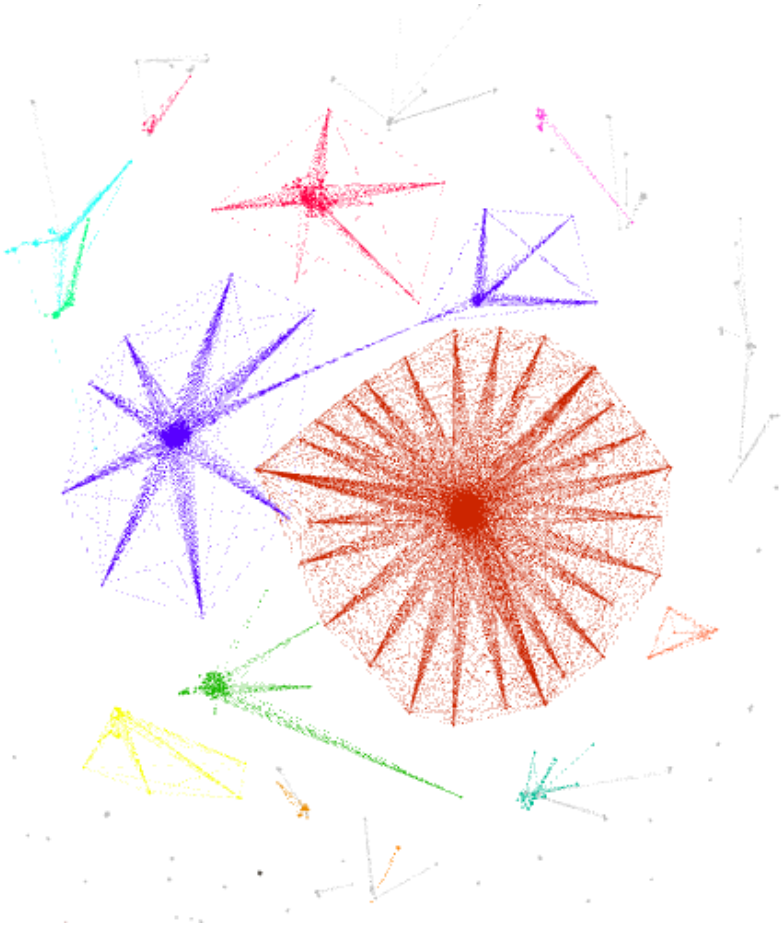
Sıra	Küme Rengi	Üye Sayısı	Küme Yüzdesi (%)
1		235	14,76
2		269	13,07
3		37	8,89
4		7	6,63
5		165	5,37
6		187	4,74
7		231	2,87
8		204	2,41
9		201	1,6
10		60	1,57
11		180	1,49

Tablo 8 B Firması Karmaşık Ağında Bulunan Kümeler ve Özellikleri

Sıra	Küme Rengi	Üye Sayısı	Küme Yüzdesi (%)
1		79	12,3
2		20	10,37
3		24	8,55
4		62	8,55
5		54	4,8
6		38	3,94
7		29	3,65
8		61	3,36
9		5	2,4
10		59	2,02
11		65	1,92



Şekil 58 A Firması Karmaşık Ağında Bulunan Kümeler



Şekil 59 B Firması Karmaşık Ağında Bulunan Kümeler

A firmasında, ana bölgeler için Tablo 7.14.'da kullanılmış olan renklerin dağılımı ve kümelenme davranışları Şekil 6.11.'de yer almaktadır. B firmasında ise, ana bölgeler için Tablo 6.15.'da kullanılmış olan renklerin dağılımı ve kümelenme davranışları Şekil 6.12.'de yer almaktadır. Veriler, satın alma verileri olması sebebiyle karmaşık ağdan çok fazla kümeler meydana gelmiştir.

6.3.3. Yapısal Analiz

Ağın düğüm sayısı, kenar sayısı, kümeleme katsayısı, yoğunluk, bağlı bileşenler, çap, yarıçap ve karakteristik yol uzunluğu ölçümleri Tablo 6.16.'da verilmiştir.

Tablo 9 A ve B Firması Karmaşık Ağının Yapısal Ölçütleri

Ölçüt Adı	A Firması - Madde Kodları Arası Ağ	B Firması - Madde Kodları Arası Ağ
Düğüm Sayısı	8840	1041
Kenar Sayısı	893508	8656
Kümelenme Katsayısı	0,931	0,846
Yoğunluk	0,023	0,016
Bağlı Bileşenler	305	90
Çap	16	7
Yarıçap	1	1
Yol Uzunluğu	3,351	2,279

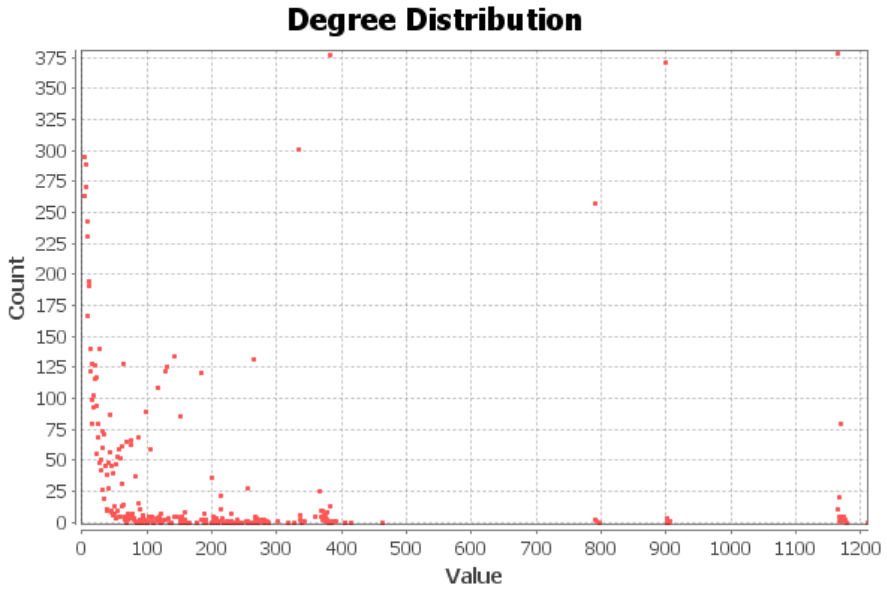
Yapısal analiz sürecinde A ve B firmaları için Tablo 6.16.'da belirtilen derece dağılım grafiği elde edilmiştir.

A firmasının, karmaşık ağ derece dağılım grafiği Şekil 6.13.'de mevcuttur. B firmasının karmaşık ağ derece dağılım grafiği ise Şekil 6.14.'de mevcuttur. A firmasında kümelenme katsayısı 0,931 olmuştur. B firmasında ise kümelenme kaysayısı 0,846 olmuştur. Yol uzunluklarında A firmasında 3,351 sonucuna ulaşılmışken, B firmasında 2,279 sonucuna ulaşılmıştır.

Degree Report

Results:

Average Degree: 202,151

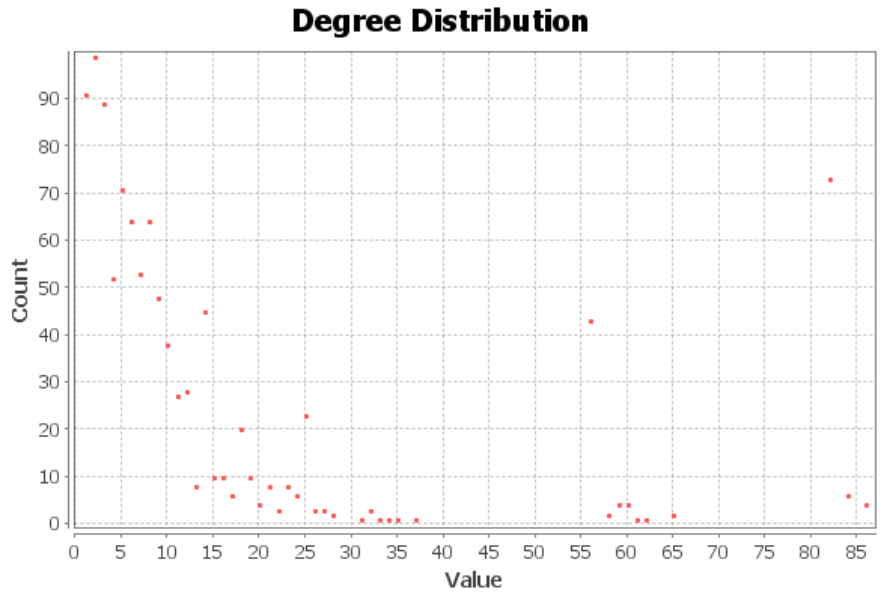


Şekil 60 A Firması Karmaşık Ağının Derece Dağılımı

Degree Report

Results:

Average Degree: 16,630



Şekil 61 B Firması Karmaşık Ağının Derece Dağılımı

Yapısal analiz sonucuna göre, Cytoscape programı kullanılarak A firması için Şekil 6.15., B firması için Şekil 6.17.'de gösterilen derece dağılım grafiği elde edilmiştir. Ağların derece dağılımını ölçekten bağımsız (scale free) gerçek hayat ağlarında bulunan, Güç Yasası Dağılımına (Power Law Distribution) benzer bir dağılımları olduğu görülmüştür [52]. Güç Yasası Dağılımı, bir ağ da derecesi çok düşük bulunan çok sayıda düğüm varken, derecesi çok yüksek bulunan çok az düğüm vardır.

A ve B firması verilerinin rastgele meydana gelen veriler olup olmadığını kontrol etmek için, Cytoscape üzerinde oluşturulan rastgele ağ ile karşılaştırmalı analizi yapılmıştır. Bu analizde network randomizer eklentisi kurularak ilgili eklenti yardımıyla A ve B firmasındaki düğüm ve kenar sayıları ile aynı düğüm ve kenar sayıları girilerek Erdös-Rényi Rassal Ağı oluşturulmuştur [53].

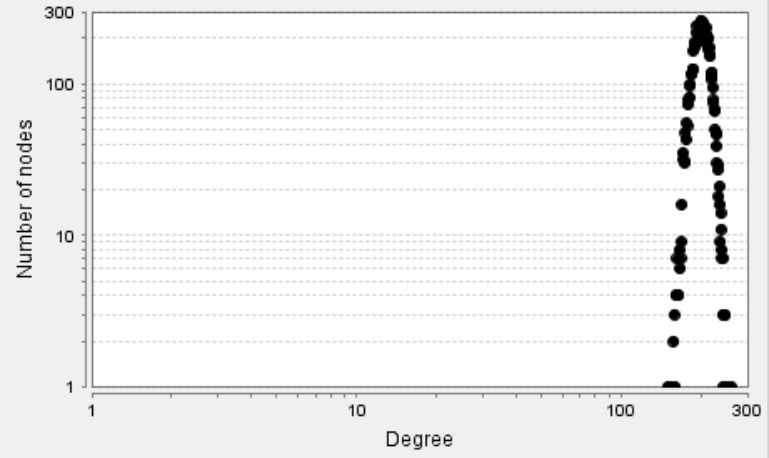
A firması için oluşturulan rassal ağın analiz sonuçları Şekil 6.16.'daki gibidir. Sonuçlara göre, kümelenme katsayısı (Clustering coefficient) 0,023 ve karakteristik yol uzunluğu (Characteristic path length) 1,987'dir.

B firması için oluşturulan rassal ağın analiz sonuçları Şekil 6.18.'deki gibidir. Sonuçlara göre, kümelenme katsayısı (Clustering coefficient) 0,016 ve karakteristik yol uzunluğu (Characteristic path length) 2,761'dir.

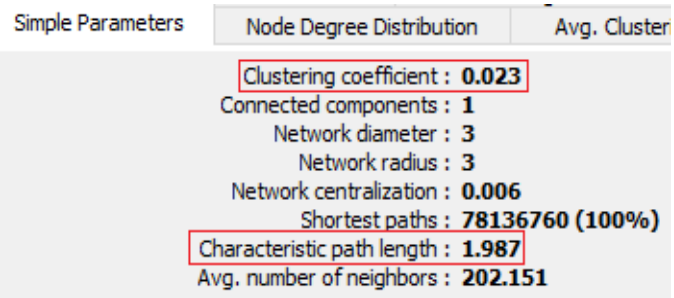
Rassal ağlar, gerçek dünya ağlarındaki yüksek kümelenme özelliğine sahip değildirler. A firması kümelenme katsayısı 0,931 değeri ile rassal ağın kümelenme katsayısı 0,023'den büyüktür. B firmasında da aynı şekilde kümelenme katsayısı 0,846 değeri ile rassal ağın kümelenme katsayısı 0,016'dan büyüktür. Karakteristik yol uzunlukları olması gerektiği şekilde rassal ağ, gerçek dünya ağlarından daha düşüktür. Buna göre, A firmasının gerçek dünya karakteristik yol uzunlukları 3,351 iken, rassal ağın karakteristik yol uzunlukları 1,987'dir. B firmasında gerçek dünya karakteristik yol uzunluğu 2,279 iken, rassal ağın karakteristik yol uzunluğu 2,761 şeklindedir. Erdös-Rényi rassal ağında, düğümler arasındaki kenarlar rastgele eklenmiş olmasından dolayı beklenildiği şekilde Poisson derece dağılımına hakimdir.

A ve B firmalarının satınalma verileri, üç temel karakteristik özellik içerisinde değerlendirildiğinde rassal ağ olmadığı, gerçek hayat ağına ait karakteristik özellikler

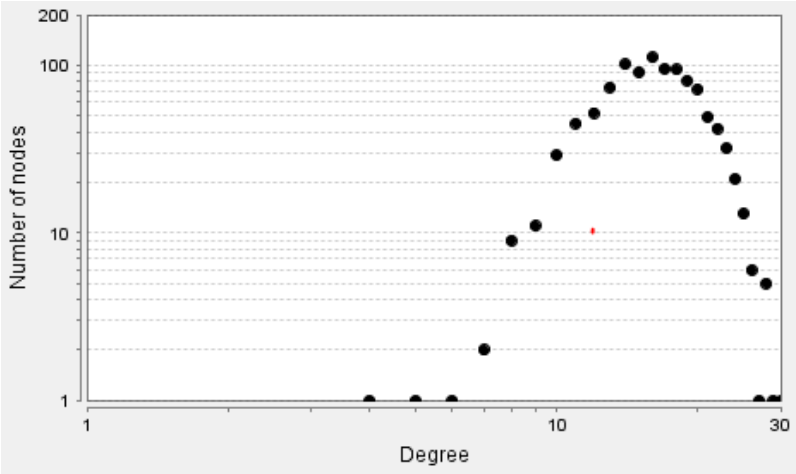
içerdiği gözlemlenmiştir. Bu analiz sonucuna göre satınalma verileri için beklenen bir sonuçtur. Sebebi ise satınalma verileri, firmaların ihtiyaçları doğrultusunda oluşmaktadır.



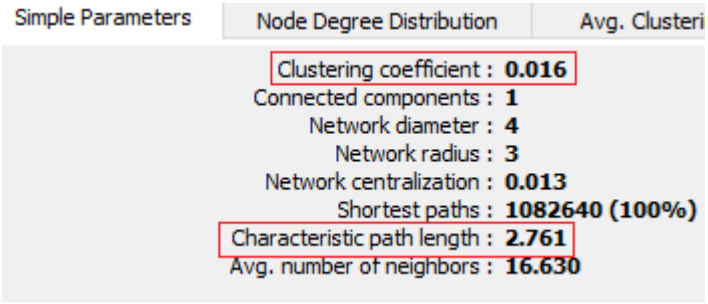
Şekil 62 A Firması Rassal Ağ Derece Dağılım Grafiği



Şekil 63 A Firması Rassal Ağ Analiz Sonuçları



Şekil 64 B Firması Rassal Ağ Derece Dağılımı Grafiği



Şekil 65 B Firması Rassal Ağ Analiz Sonuçları

BÖLÜM 7. BULGULAR VE YORUMLAR

İki firmanın satınalma verilerinden yapılan birliktelik analizinde, birliktelik kümeleri araştırılarak madde kodlarının kendi içlerindeki destek ve güven birliktelikleri gözden geçirilmiştir. Yapılan çalışma sonrası satın almalarda en fazla 3'lü birlikteliklerin olduğu görülmüştür. Bu çalışma neticesinde satınalma birimlerine aşağıdaki önerilirde bulunulabilir;

- 3'lü ve 2'li birlikteliklerdeki madde kodlu ürünler alınırken diğerininde aynı şekilde kullanılacağı düşünülerek diğer madde kodundaki ürünlerinde alınması önerilebilir.

- En çok alınmış olan madde kodundaki ürünlerden daha fazla stok yapılabilir. Bu sayede ihtiyaç duyulduğunda satınalma sürecini beklemeye gerek kalmayacak ve bu süreçte ürünlere zam gelmesi durumunda zamdan etkilenilmeyecektir.

- Birlikte alınan ürün setlerinde tedarikçi firmalar tarafından uygun indirimler istenebilir.

- En az alınan ürünlerde stok yapılmayabilir veya stok sayısı az olacak şekilde ayarlanabilir.

- En çok satınalınan ürünler depolarda daha ön raflara alınabilir ve daha kolay ulaşılabileceği raflara dizilebilir.

- En çok satın alınan ürünler incelenebilir. İnceleme sonucu rakamların yüksek görülmesi halinde nedeni araştırılabilir ve buna göre ilgili ürünlerin satınalımları düşürülebilir. Örnek; Elektrik prizlerinin çok alındığının görülmesi halinde, prizlerin kırılma veya bozulma sonrası mı alındığı yoksa yeni bir ofis yapımı sırasında kullanılmak üzere mi alındığı araştırılabilir. Satın alımların prizlerin kırılması nedeniyle ise şirket çalışanları bu konuda uyarılabilir. Kırılma ve bozulmaların azaltılmasıyla şirkete kâr sağlanabilir.

Destek ve güven birlikteliği olan madde kodlarında, hangi ürünler için stok yapılabileceği konusunda fikir verebilir sonuçlara ulaşılmıştır. Örneğin; güven değeri %100 olan madde kodu ve madde kodu setleri olduğu görülmüştür.

İki firma için yapılan karmaşık ağ analizinde, her grup ürünün kendi içerisinde küme oluşturduğu ve bu yüzden çok fazla kümelerin meydana geldiği görülmüştür. Bunlara örnek; Cıvata ile pullar ve rulmanlar gibi ürünler kendi içlerinde kümeler oluşturmuştur. Bu çalışma neticesinde yine birliktelik analizindeki gibi oluşan kümelere göre raf dizilimi ve stoklama yapılabilir.

Yapılan birliktelik analizi ve karmaşık ağ analizi neticesinde çıkarılmış olan sonuç, satınalma açısından faydalı olabilecek bulgulara ulaşılmış olmasıdır. Bu bulgular doğrultusunda çeşitli aksiyonlar önerilmiştir. Yapılan öneriler doğrultusunda firmaya kârlılık açısından faydalı sonuçlara ulaşılmıştır. Aynı zamanda doğru stoklama ve raf diziliminin daha doğru şekilde yapılması konusunda öneriler sunulmuştur.

BÖLÜM 8. SONUÇ

Bu çalışma bir holding bünyesine bağlı bulunan, Türkiye'nin önde gelen firmalarının verileri üzerinde gerçekleştirilmiştir. Satınalma verileri üzerinden birliktelik kuralları ve ağ analizleri yapılarak incelenmiştir. Birliktelik kuralları ve ağ analizi yapılmış olan sektörler, bu analizler sonucunda satınalma stratejilerini farklılaştırabilirler. Bu çalışma sayesinde satın alımlarda daha çok hangi ürünlerin alındığı konusunda fikir veren sonuçlar ortaya çıkmıştır. Ayrıca matematiksel ve standart yöntemler ile bulunamayan sonuçlar, Apriori Algoritmasıyla bir takım tespitlere ulaşılmıştır.

Birliktelik analizi, kullanılan verilerin temizlenip, düzenlenmesi sonrası Apriori Algoritması ile yapılmıştır. Birliktelik analizi için KNIME uygulaması kullanılmıştır. Kullanılan verinin gelişimi ise KNIME uygulaması içerisine node'ların eklenmesiyle yapılmıştır.

Birliktelik analizi neticesinde çıkan sonuçlar, ilgili sektörler açısından faydalı olabilecek bir takım bilgilere ulaşmıştır. Yapılan bu çalışma sonucunda stoklamada ve rafların doğru dizilişinde faydalı olacağı ön görülmüştür.

Karmaşık Ağ Analizi için SQL yardımı ile firma verileri ağ analizine uygun hale getirilmiştir. Ağ analizde Gephi ve Cytoscape programı kullanılarak ürünler arası ağlar ortaya çıkarılmıştır. Gephi programı ile derece, yakınlık, arasındalık ve özvektör merkezilik analizleri yapılmıştır.

A ve B firmalarının satınalma verilerinde yapılan görsel ve topluluk analizlerinde kümelenme davranışı gösterdiği gözlemlenmiştir.

Karmaşık ağ analizinde son olarak yapılan yapısal analizinde ise A ve B firması satınalma verilerinin gerçek dünya ağlarına ait karakteristik özelliklere sahip olup olmadığı incelenmiştir. Bu analizde aynı düğüm ve kenar sayısı kullanılarak Erdös-Rényi rassal ağı meydana getirilmiş ve gerçek dünya ağları ile rassal ağ sonuçları karşılaştırılmıştır. Çıkan sonuçlara göre, A ve B firması verilerinin gerçek hayat ağının karakteristik özelliklerine sahip olduğu gözlemlenmiştir.

Yapılmış olan iki çalışma neticesinde, birliktelik analizinde destek ve güven değerlerinde en fazla satın alınmış olan madde kodlarının, karmaşık ağ analizi merkezilik sonuçlarındaki değerleri incelenmiştir. İncelenmiş olan değerlere göre birliktelik analizinde ön plana çıkmış olan madde kodlarının, merkezilik değerlerinin düşük olduğu gözlemlenmiştir.

Verilerin büyük bir market verileri olmaması ve farklı sektörlerden veriler olması sebebiyle, market veri sonuçlarındaki gibi yüksek değerler vermediği gözlemlenmiştir.

İncelenmiş olan verilerin birliktelik ve karmaşık ağ analizinde önemli sonuçlara varılmıştır. Bu analizlerin çıkan sonuçlarına göre, kullanılmış olan veride birlikte satın alınan en fazla 3'lü ürün kümesi olduğu görülmüştür. Destek ve güven değerlerine göre de analizler yapılmıştır. Güven değeri %100 olan ürünlerin birliktelikleri olduğu sonucuna ulaşılmıştır. Diğer bir değişle %100 güven katsayılarında olan ürün setlerinin şimdiye kadar hep birlikte satın alındığı görülmüştür. Karmaşık ağ analizinde ise her grup ürünün kendi içerisinde küme oluşturduğu ve bu yüzden çok fazla kümelerin meydana geldiği görülmüştür. Bu ürünler stoklama yapılırken daha çok dikkate alınabilir, tedarik sürecinde beklememek ve üretimin aksamaması adına bu ürünlerden daha fazla stoklama yapılması önerilebilir. Raf dizilimi yapılırken bunlara daha çok yer ayrılabilir ve ön sıralara alınabilir. Yine bu ürünlerin toplu alımları sırasında tedarikçi firma tarafından topla alım sebebiyle daha fazla indirim istenebilir veya şirket içerisinde kullanılmakta olan farklı ürünlerden hediye edilmesi istenebilir.

Bu çalışmada yapılan analizler madde kodu bazında gerçekleştirilmiştir. Buna göre: ileride bu analizleri farklı bir çalışmada madde kodundaki ürünleri kategorize ederek, kategori bazlı olarak çalışma yapılabilir. Kategori bazlı çalışmada elektrik ürünleri, mutfak ürünleri, idari işler için olan ürünler gibi farklı gruplar çıkarılarak analiz yapılabilir. Aynı şekilde analizlere farklı parametreler eklenerek de bu çalışma yapılabilir. Örneğin, yangın sonrası günlerde alınan ürünler, ekonomik kriz sırasında alınan ürünlerin birliktelikleri çıkarılarak farklı stratejiler uygulanabilir.

KAYNAKÇA

[1] S. Özekes, "Veri Madenciliği Modelleri ve Uygulama Alanları," *İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü Dergisi,* vol. 2, no. 3, 2003.

[2] C. Mauri, "Card loyalty. A newemerging issue in grocery retailing," *Journal of Retailing and Consumer Services,* vol. 10, no. 1, pp. 13-25, 2003.

[3] T. Raeder and N. V. Chawla, "Market basket analysis with networks," *Social Network Analysis and Mining,* vol. 1, pp. 97-113, 2011.

[4] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in very large databases," presented at the In Proceedings of the 20th International Conference on VLDB, 1994.

[5] G.Adomavicius and A.Tuzhilin, "User profiling in personalization applications through rule discovery and validation," presented at the KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, New York/USA, 1999.

[6] V. Tunalı and E. Güvenoğlu, "Complex Network Analysis Approach to Examining Undergraduate Program Preferences," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi,* vol. 8, no. 1, pp. 176-186, 2019.

[7] H. K. Kim, J. K. Kim, and Q. Y. Chen, "A product network analysis for extending the market basket analysis," *Expert Systems with Applications,* vol. 39, no. 8, pp. 7403-7410, 2012.

[8] Y. Demirok, "Birliktelik Kuralları Yöntemi İle E-Ticaret Satışlarının Analizi," Yüksek Lisans, Bilgisayar Mühendisliği, Maltepe Üniversitesi, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı 2018.

[9] D. Çelik, "Veri Madenciliği Kullanılarak Akıllı Reklam/Anket Uygulaması," Yüksek Lisans, Bilgisayar Mühendisliği, İstanbul Üniversitesi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 2015.

[10] I. F. Videla-Cavieres and S. R. Rios, "Extending market basket analysis with graph mining techniques," *Expert Systems with Applications,* vol. 41, no. 4, pp. 1928-1936, 2014.

[11] O. Y. Bolata, "Ülkeler Arası Göç Ağlarının Modellenmesi ve Analizi," Yüksek Lisans, Bilgisayar Mühendisliği, Maltepe Üniversitesi, Maltepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 2018.

[12] A. Vahaplar and D. M. M. İnceoğlu, *Veri Madenciliği ve Elektronik Ticaret*. Ege Üniversitesi Bilgisayar Mühendisliği Bölümü, 2001.

[13] G. Acun, "Yazılım Hata Logları Kullanılarak Veri Madenciliği Uygulaması Gerçekleştirilmesi," Yüksek Lisans, Bilgisayar Mühendisliği, Maltepe Üniversitesi, Maltepe Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı, 2014.

[14] T. B., "Data Mining for Malicious Code Detection and Security Applications," presented at the European Intelligence and Security Informatics Conference, Milan / Italy, 2009.

[15] Y. Özkan, *Veri Madenciliği Yöntemleri*. Papatya Yayıncılık, 2013.

[16] M. F. Alaeddinoğlu, T. Aydın, and D. Dal, "Birliktelik Kuralları İle Mekansal Zamansal Veri Madenciliği," *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi,* no. 5, pp. 191-212, 2012.

[17] M. Zerman, "Birlitelik Kuralları Algoritmaları ile Büyük Veriler Üzerinde Analitik Analizler: Havaalanı Örneği," Yüksek Lisans, Bilgisayar Mühendisliği, Haliç Üniversitesi, Haliç Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı, 2018.

[18] M. Ö. Dolgun, "Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının Bağımlı Değişken Prevelansı Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası Ilişki Yapısına Göre Karşılaştırılması," Biyoistatistik Programı, Hacettepe Üniversitesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, 2014.

[19] N. Z. a. L. Zhou, "Methodologies for Knowledge Discovery and Data Mining " presented at the Third Pacific-Asia Conference, PAKDD-99, Beijing, China, 1999.

[20] H. Akpınar, "Veri Tabanlarında Bilgi Kesfi ve Veri Madenciligi," *İstanbul Üniversitesi İşletme Fakültesi Dergisi,* 2000.

[21] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining Association Rules between Sets of Items in Large Databases," presented at the International Conference on Management of Data, Washington DC, 1993.

[22] R. Agrawal and R. Srikant, "Mining Sequential Patterns," presented at the Published in Proceedings of the Eleventh International Conference on Data Engineering, San Jose 1995.

[23] W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge discovery in databases: An overview," *AI Magazine,* vol. 13, no. 13, 1992.

[24] H. Zhu, "On-Line Analytical Mining of Association Rules," Master Degree, Simon Fraser University, 1998.

[25] B. Doğan, B. Erol, and A. Buldu, "Sigortacılık Sektöründe Müşteri İlişkileri Yönetimi İçin Birliktelik Kuralı Kullanılması," *Marmara Fen Bilimleri Dergisi,* vol. 3, pp. 105-114, 2014.

[26] M. Houtsma and A. Swami, "Set-Oriented Mining for Association Rules in Relational Databases," presented at the Proceedings of the 11th IEEE International Conference on Data Engineering, Taiwan, 1995.

[27] A. Das, W.-K. Ng, and Y.-K. Woon, "Rapid Association Rule Mining," presented at the In Proceedings of the Tenth International Conference on Information and Knowledge Management, Atlanta, GA, USA,, 2001.

[28] M. J. Zaki and C.-J. Hsiao, "CHARM: An Efficient Algorithm for Closed Itemset Mining," presented at the In 2nd SIAM International Conference on Data Mining, 2002.

[29] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data," *Knowledge Engineering and Data Science,* vol. 1, no. 1, pp. 20-25, 2017.

[30] J. Singh, H. Ram, and J. S. Sodhi, "Improving Efficiency of Apriori Algorithm Using Transaction Reduction," *International Journal of Scientific and Research Publications,* vol. 3, no. 1, pp. 1-4, 2013.

[31] T. Özseven and M. Düğenci, "LOG Analiz: Erişim Kayıt Dosyaları Analiz Yazılımı ve GOP Üniversitesi Uygulaması " *Bilişim Teknolojileri Dergisi,* vol. 4, no. 2, pp. 55-66, 2011.

[32] J. Han and M. Kamber, *Data Mining Concept and Technology* (Morgan Kaufmann series in data management systems.). San Francisco : Morgan Kaufmann Publishers, 2001, pp. 70–72.

[33] N. Gürsakal, *Sosyal Ağ Analizi*. Dora Yayınları, 2009.

[34] L. Euler, "Solutio problematis ad geometriam situs pertinentis," *Acad. Scient. Petropol. ,* vol. 8, pp. 128-140, 1741.

[35] J. Scott, *Social Network Analysis A Handbook*. Sage, 2000.

[36] B. Akkoç, "Sosyal Ağ Analizi için Bayes Ağlarının Kullanımı," Yüksek Lisans Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Selçuk Üniversitesi, Selçuk Üniversitesi, 2012.

[37] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and E. Bhattacharjee, "Measurement and analysis of online social networks," presented at the Proceedings of the 7th ACM SIGCOMM Conference on Internet measurement, San Diego, California, USA, 2007.

[38] L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004.

[39] H. Gönültaş, "E-Posta Listelerinde Metin Kümeleme ve Sosyal Ağ Analizi Uyumu," Yüksek Lisans, Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Gebze Yüksek Teknoloji Mühendislik ve Fen Bilimleri Enstitüsü, 2010.

[40] K. H. Rosen, *Discrete Mathematics and its Applications*. WCB/McGraw-Hill, 1999.

[41] P. Mika, *Social Networks and the Semantic Web*. New York: Springer Science and Business Media, 2007.

[42] M. Gencer, C. Gunduz, and V. S. Tunalioglu, "CL-SNA: Social Network Analysis with Lisp," presented at the In Proceedings of the 2007 International Lisp Conference on - ILC '07, New York, USA, 2007.

[43] S. Wasserman and J. Galaskiewicz, *Advances in Social Network Analysis : Research in the Social and Behavioral Sciences*. California: Sage Focus Editions, 1994.

[44] H. Kuduğ, "Sosyal Ağ Analizi Ölçütlerinin İş Ağlarına Uyarlanması," Yüksek Lisans, Bilgisayar Mühendisliği, Ege Üniversitesi, Ege Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 2011.

[45] M. E. J. Newman, *Mathematics of Networks*. England: Palgrave Macmillan, 2008.

[46] V. Tunalı, *Sosyal Ağ Analizine Giriş*. Ankara: Nobel Akademik Yayıncılık, 2016.

[47] T. T. Bilgin, "Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları," presented at the Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri, Şanlıurfa, 2009.

[48] K. Gazioğlu and Ş. E. Şeker, "Veri Madenciliği Yöntemleri İle Twitter Üzerinden Girişimcilik Analizi," *YBS Ansiklopedi,* vol. 4, no. 4, 2017.

[49] A. Mihanović, H. Gabelica, and Ž. Krstić, "Big Data and Sentiment Analysis using KNIME: Online Reviews vs. Social Media," presented at the 37th International Convention on. IEEE, 2014.

[50] https://gephi.org/about/. (2020). *Gephi*.

[51]https://www.wikizeroo.org/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvR2VwaGkjY2l0ZV9ub3RlLVVzaW5lTm91dmVsbGUtR2VwaGktNA. (2020). *Gephi Tarih*.

[52] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.,* vol. 74, no. 1, pp. 47–97, 2002.

[53] P. Erdös and A. Rényi, "On random graph I," *Publ. Math.,* vol. 6, pp. 290–297, 1959.