Dicle University Journal of Engineering (DUJE)
web: http://dergipark.gov.tr/dumf

Araştırma Makalesi / Research Article

Çoklu Sosyal Ağlarda Aynı Kullanıcıları Belirleme Yöntemi

Method to Determine the Same Users on Multiple Social Networks

Ahmet Müngen[1], Betül Bulut[2], Mehmet Kaya[3*]

[1]Ostim Teknik Üniversitesi, Yazılım Mühendisliği Bölümü, Ankara, amungen@gmail.com
[2]Tarsus Üniversitesi, Bilgisayar Teknolojileri Bölümü, Mersin, betulbulut@tarsus.edu.tr
[3]Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, Elazığ, kaya@firat.edu.tr

MAKALE BİLGİLERİ

Anahtar kelimeler:

çoklu sosyal ağlar, düğüm hizalama, vektör tabanlı benzerlik algoritması, aynı kullanıcıların tespit edilmesi

Makale geçmişi:

Geliş: 1 Temmuz 2020
Düzeltme: 3 Temmuz 2020
Kabul: 13 Temmuz 2020

ÖZET

Kullanıcılar tarafından değişik amaçla kullanılan her sosyal ağ farklı kullanıcı verileri içermektedir. Kullanıcıların farklı sosyal ağlardaki hesaplarını bulmak ve bulunan verileri birleştirip tek bir veri havuzunda derlemek hem tavsiye edici sistemleri iyileştirecek hem de kullanıcı deneyimini attıracak çok önemli bir etken olacaktır. Çalışma kapsamında özgün düğüm hizalama ve düğüm benzerlik yöntemleri önerilmiştir. Topolojik bazlı düğüm önermede çapa yöntemi kullanılırken bağlantılar arası yoğunluk ilişkileri de dikkate alınmıştır. Benzerlik tabanlı düğüm benzerlik yönteminde ise öznitelik seçim kriterleri, başlangıç noktası tespit problemi ve değişken formülasyon ile başarılı düğüm eşleştirme sayısı arttırılmıştır. Bununla birlikte bu çalışmada hem kullanıcıların profil özelliklerine göre hem de diğer kullanıcılar ile aralarındaki ilişkilere göre hizalama ve benzerlik tespiti yapılmıştır. Farklı sosyal ağlardaki aynı hesaplarının bulunması ile ilgili dokuz farklı metot önerilmiştir. Önerilen yöntemler İki ile altı arasında değişen sosyal ağ verilerinde kadar toplanan sosyal ağlarda test edilmiş ve kullanıcıların eşleşme başarı oranları ölçülmüştür. Bu sonuçlarda, %95'lere varan başarı oranları yakalanmıştır. Böylece çoklu sosyal ağlarda farklı öznitelikleri aynı graf üzerinde toplanmış kullanıcılar için birden fazla sosyal ağı kapsayan geniş bir kullanıcı profili oluşturulmasına imkân sağlanmıştır.

Doi: 10.24012/dumf.761949

ARTICLE INFO

Article history:

Received: 1 July 2020
Revised: 3 July 2020
Accepted: 13 July 2020

Keywords:

multiple social networks, node alignment, vector-based similarity algorithm, identifying the same users

ABSTRACT

Each social network used by users for different purposes contains different user data. Finding users' accounts in different social networks and combining the data found and compiling them into a single repository will be a very important factor that will both improve the recommended systems and increase the user experience. Within the scope of the study, original node alignment and node similarity methods are proposed. While using the anchor method in topological-based node proposition, density relationships between the links are also taken into consideration. In the similarity based node similarity method, attribute selection criteria, starting point detection problem and variable formulation have increased the number of successful node matching. However, in this study, alignment and similarity were determined both according to the profile characteristics of the users and the relationships between them. Nine different methods have been proposed to find the same accounts in different social networks. Proposed methods Tested on social networks collected in social network data ranging from two to six, and match success rates of users were measured. In these results, success rates up to 95% have been achieved. Thus, it is possible to create a wide user profile covering multiple social networks for users whose different attributes are gathered on the same graph in multiple social networks.

* Sorumlu yazar / Correspondence
Mehmet Kaya
✉ kaya@firat.edu.tr

Please cite this article in press as A.Müngen, B.Bulut, M.Kaya, "Çoklu Sosyal Ağlarda Aynı Kullanıcıları Bleirleme Yöntemi", DUJE, vol. 11, no. 3, pp. 1043-1054, September 2020.

Giriş

Platformlar arası sosyal ağ birleştirmenin ana fikri; farklı sosyal ağlardaki kullanıcıların hesaplarının tespit edilmesidir. Bu konunun araştırılması, platformlar arası sosyal ağ araştırması için bir temel oluşturacaktır. Hiçbir çözüm tüm anonim sosyal ağ kullanıcılarını tanımlayamaz. Ancak bazı sosyal ağ öznitelikleri, birden fazla sosyal ağdaki kullanıcıların bir kısmını tanımlamak için kullanılabilir. Kullanıcıların özniteliklerine erişmek için özel bir izne gerek yoktur ve bu öznitelikler kullanıcılar tarafından farklı amaçlarla (kötü niyetli kullanıcılar dahil) kolayca taklit edilebilir. Bu nedenle detaylı analiz edilmeyen düğüm benzerlik yöntemleri oldukça güvensizdir. Konum verilerinin sosyal ağlardan elde edilmesi ve yazma stilinin kısa cümlelerden çıkarılması zor olduğundan düğüm benzerlik bulma algoritmalarında GPS ve yazım türü gibi veriler pek kullanılmaz. Her ne kadar kullanıcıların bağlantıları toplanabilse de ve neredeyse tüm sosyal ağlarda taklit profiller ile bir kullanıcının taklit edilmesi mümkün olsa da, taklit unsuru dışarıda tutularak kullanıcıları tanımlamak için yapılan bazı çalışmalara rastlanmıştır [1].

Sosyal ağlar tasarlanırken kullanıcıların en çok işlerine yarayacak özellikler, öznitelik olarak belirlenir. Her sosyal ağın ihtiyacı, hedef kitlesi ve öz niteliklerinin farklı olması beklenen bir durumdur. Sadece fotoğraf paylaşımı yapılan sosyal ağlarda bile fotoğrafın paylaşım amacının değişmesi özniteliklerin değişmesine yol açar. Örneğin fotoğrafın hangi makine ile çekildiği, nerede çekildiği ve diğer profesyonel fotoğrafçıları ilgilendiren detayların bulunduğu bir sosyal ağın öncelikleri, hızlı paylaşım yapılan, filtre eklenen ve etiket modeli ile çalışan bir sosyal ağın önceliklerinden farklıdır. Her sosyal ağ, kullanıcılarının sağladığı bilgiler ve kullanıcıların birbiriyle iletişiminden ötürü ayrı bir uzay olarak ifade edilir. Her sosyal ağ ayrı bir uzay olarak düşünüldüğünde tamamen birbirinden farklı heterojen uzaylar bütünlüğü oluşur. Kullanıcılar farklı uzaylarda farklı bilgileri ve ürettikleri farklı eserleri paylaştığı için farklı sosyal veriler genelde özgün bilgi içerir. Bu çalışmada farklı sosyal ağlarda yani farklı uzaylarda profili olan kullanıcıların profil bilgilerinin birleştirilmesi amaçlanmıştır. Farklı sosyal ağların birleştirilmesi son yıllarda popülaritesi artan bir konudur. Bu alanda yapılmış çalışmalar ve yayınlanmış bazı yöntemler mevcuttur [2]. Bu konuda yapılan çalışmalar genel olarak ikiye ayrılır: Birincisi özniteliklerin ilişkisine bakarak farklı ağlardaki düğümlerin benzerlik bazında eşleştirilmesidir. İkincisi ise uzay üzerindeki düğümler arasındaki ilişkiyi analiz ederek yapılan düğün birleştirme işlemidir.

Çalışmanın bundan sonraki bölümleri aşağıdaki gibidir: Literatürde mevcut yöntemler bölümünde düğüm benzerlik ve topolojik hizalama yöntemlerinden bahsedilmiş ve literatürde yapılan çalışmalara yer verilmiştir. Veri Toplama ve Sosyal Ağ seçimi bölümünde veri kümelerinin toplanması ile ilgili işlemler anlatılmıştır. Sonraki bölümlerde önerilen yöntemler ve elde edilen sonuçlar anlatılmıştır. Son bölümde ise sonuç bölümüdür.

Literatürdeki Mevcut Yöntemler

Düğüm Benzerlik Yöntemleri

Düğüm benzerlik yöntemi ile ilgili iki temel yaklaşım vardır. İlk yaklaşım, esas olarak düğüm yapısına odaklanan bir ağın yerel özelliklerine dayanır; ikinci yaklaşım, bir ağdaki genel yol yapısını tespit eden global özelliklere dayanmaktadır.

Bir ağdaki düğümlerin "benzerliği" analiz etmek için en bilinen yerel benzerlik yöntemleri Adamic Adar [3], Jaccard [4] ve Ortak Komşular [5] endeksidir. Bu yöntemler arasında Adamic Adar [3] yönteminin bir sosyal ağdaki yeni bağlantıları tahmin etmede en iyi performansı gösterdiği bildirilmiştir. Jaccard Benzerlik yöntemine benzeyen Adamic Adar ilk olarak iki web sayfasının ne kadar benzer olduğunu ölçmüştür. Friend of Friend algoritması (FOAF) olarak da bilinen Ortak Komşular yöntemi ise [5], arkadaş tavsiyesi görevi için facebook.com ve hi5.com gibi birçok popüler sosyal ağ tarafından benimsenmiştir. FOAF, birçok ortak komşuları olan iki düğümün gelecekte bir bağlantı

oluşturma olasılığının yaygın olduğu fikrine dayanır. Ayrıca, diğer yerel benzerlik önlemleri tercihli bağlılığa dayanmaktadır. Tercihli bağlamanın temel dayanağı, yeni bir kenarın bir düğümü içermesi olasılığının, komşularının mevcut sayısıyla orantılı olmasıdır. Bu konuda çeşitli küresel yaklaşımlar vardır. Bunlardan bazıları En Kısa Yol algoritması, RWR algoritması ve SimRank algoritmasıdır. Liben ve Kleinberg [6] bir grafikteki herhangi bir düğüm çifti arasındaki en kısa yolun tanımlanmasının bağlantı tahmini için kullanılabileceğini iddia etmişlerdir. İki düğüm arasındaki en kısa yolun hesaplanması, bilinen herhangi bir en kısa yol algoritması kullanılarak yapılabilir [7]. RWR algoritması [8] (Yeniden Başlatma algoritması ile Rastgele Yürüme) bir grafik boyunca rastgele yürüyüşün Markov zincir modeline dayanmaktadır. RWR ile aynı yönde Fouss ve ark. [9], bir ağdaki herhangi bir düğüm çifti arasındaki benzerlikleri elde etmek için bir Rastgele bir yürüyüş modeli önermişlerdir. Bu düğümler, iki düğümü bağlayan yolların sayısı arttığında ve bunları bağlayan yolların uzunluğu azaldığında artar. Symeonidis ve Tiakas [10] önerdikleri FriendTNS yöntemi ile hem zaman merkezli hem de başarı oranı yüksek olan bir benzerlik yöntemi önermişlerdir. SimRank [11] bir ağın yapısal bağlamına dayanan küresel bir benzerlik ölçüsü de hesaplar. Son zamanlarda Clauset ve ark. [12] hiyerarşik ağ yapısına dayanan bir algoritma önermişlerdir. İlk olarak, gerçek ağ verilerine istatistiksel olarak sığdırmak için hiyerarşik rastgele bir grafik kullanırlar. Daha sonra, bağlantı olasılığının hiyerarşideki düğümlerin derinliğine bağımlılığı çıkarılabilir. Ağın eksik bağlantıları, bağlantı olasılığına göre azalan sırada sıralanarak tahmin edilebilir. Son olarak, Blondel ve ark. [6] farklı yönlendirilmiş ağların düğümleri arasındaki yol temelli benzerlik ölçümlerini dikkate almıştır [13].

Düğüm Hizalama Yöntemleri

Çevrimiçi sosyal ağlarda kullanıcıları bir graf veya vektörel bir uzay üzerinde haritalamak son zamanlarda akademik çalışmalarda sık görülmektedir [14]. Kullanıcı etkileşimleri ile oluşturulan harita sayesinde bağlantı tahmini gibi uygulamalardan yararlanabilir [15] ve öneri sistmleri için kullanılabilir [16]. Heterojen sosyal ağlarını hizalama, veri seyrekliği sorununu hafifletebilir ve sosyal ağ analizinde yararlı çalışmalar yapmak için önemlidir. Ayrıca kullanıcıların hesap bilgilerini genelde değiştirmedikleri söylenebilir [17]. Ayrıca, farklı ağlardaki demografik bilgilerin dengesiz olma olasılığı yüksektir ve zengin ve doğru profiller ile her zaman karşılaşılmayabilir. Bazı araştırmacılar dil stilini ve ifadeyi benimsemeyi kullanarak sınıflandırmayı önerdiler [18]. Alternatif olarak, sosyal ağların yapısal bilgileri doğrudan kullanıcı uyumu için kullanılabilir. Ağlar arası, bağlantılar ve ortak kullanıcılar olasılıksal bir grafik sınıflandırıcı elde etmek için kullanılabilir [19] veya alaka düzeyi hesaplaması için birden fazla ağın ortak bir alt uzayı oluşturulabilir [20]. Bu ilgili çalışmaların çoğu bağlantıların yönlendirilmemiş olduğunu düşünmektedir. Müngen ve arkadaşları önerdikleri çalışmada hem düğüm bilgileri hem de düğümler arası ilişkilere bakarak farklı ağlardaki aynı kişiye ait hesapları bulmaya çalışan öncü bir çalışma yapmışlardır [21]. Zhang ve arkadaşları topoloji tabanlı hizalama sürecine rehberlik etmek için düğüm nitelik bilgisinden faydalanan bir düğüm hizalama yöntemi önermişlerdir [22]. Trung ve arkadaşları gömme modeline dayanan tamamen denetlenmeyen bir ağ hizalama çerçevesi önermiştir [23]. Zhan ve arkadaşları birleşik bir bağlantı tahmini çerçevesi önererek toplu bağlantı füzyonunu (CLF) temel alan kolektif rastgele yürüyüş ile kısmen hizalanmış olasılıksal ağlar kullanmışlardır [24].

Veri Toplama ve Sosyal Ağ Seçimi

Sosyal Ağlar Karşılaştırılması

Çalışmanın seçilen veri setlerinden 3 adet sosyal ağ metin tabanlı, 3 adet sosyal ağ fotoğraf tabanlı, 2 adet sosyal ağ coğrafya tabanlı ve 1 adet sosyal ağ video tabanlıdır. Metin tabanlı olmayan sosyal ağlarda da analizlerin birçoğu

diğer paylaşım tipleri ile birlikte metin olarak sunulan veriler kullanılarak yapılmıştır. Örneğin fotoğraf bazlı paylaşımı ön plana çıkaran Instagram sosyal ağında fotoğrafın açıklaması ve etiketlerini gösteren metin ele alınmıştır.

Tüm sosyal ağlar kullanıcılarını tanımlamak için tekil bir kullanıcı adı alınmasını zorunlu tutmaktadır. Meetup ve Foursquare sosyal ağlarında lokasyon alma ve lokasyonun doğru tespit başarı oranları coğrafi tabanlı sosyal ağlar olduğu için çok yüksektir. Kullanılan dillerin tespiti, sosyal ağda paylaşılan içerikler ve varsa kullanıcının kendini ifade ettiği açıklamalardan alınarak yapılmaya çalışılmıştır. Çok dil kullanan kullanıcıların görülmesi sosyal ağlarda normaldir. Bir profil için İngilizce ve İspanyolca dışında kalan dilin ağırlığı bu iki dilin ağırlığının iki katından fazlaysa kullanıcı yerel dili olarak atanmaktadır. Aktif Saat Dilimi kullanıcının sosyal ağı kullandığı aktif saat dilimini vermektedir. Yirmi dört saatten oluşan gün 6 eşit parçaya bölünerek kullanıcının hangi parçada sosyal ağda daha aktif olduğu tespit edilmeye çalışılmıştır. Diğer sosyal ağ bağlantıları kullanıcıların beyan ettikleri diğer sosyal ağların linklerini / listesini içerir. Ortak etkinlik kaydı kullanıcıların birlikte katıldıkları veya katılmak istediklerini beyan ettikleri etkinliklerin kaydını içerir.

Tablo 1. Etkileşim Türü Olarak Tablo

Özellik / SA	T	I	F	M	L	P	R	F	Y
Beğeni (Like)	✓	✓	✓	-	✓	✓	✓		✓
Tekrarlama (Retweet)	✓	T	-	-	-	T	-	-	-
Yorum	✓	✓	✓	-	✓	✓	✓	-	✓
Bahsetme (Mention)	✓	✓	✓	-	-	✓	-	-	✓

Tablo 1'de etkileşim türlerine göre sosyal ağlar verilmiştir

Tablo 2. Lokasyon ve Dil Tespiti

Özellik / SA	T	I	F	L	P	R	Y
Açıklamada Lokasyon Bilgisi	✓	✓	✓	✓	-	✓	-
Paylaşılan İçerik	✓	✓	✓	✓	-	✓	✓
Yorum	✓	✓	✓	✓	✓	✓	-

Tablo 2'de lokasyon ve dil tespiti bazlı sosyal ağ sınıflandırması sunulmuştur.

Sosyal ağlar, Tablo 1'de ifade edilen özelliklerden çok daha geniş bilgiler verebilir. Birçok sosyal ağ kullanıcısı tüm verileri doldurmamakta veya halka açık olarak sunmamaktadır. Bundan dolayı bu çalışmanın veri toplama aşaması hazırlanırken, profillerin en az yarısını dolduran kullanıcıların verileri ele alınmıştır.

Ortak Veri Havuzu Hazırlama

Çalışmada birden fazla sosyal ağ hesabına sahip kullanıcıların aynı kişi olup olmadıklarının tespiti üzerine olacağı için sosyal ağlardan elde edilen verilerde aynı kullanıcıların olması gerekmektedir.

Çalışmamızda çapa yöntemi kullanıldığı için hali hazırda en az bir düğümün birden fazla sosyal ağda birbiri ile başlangıç durumu olarak eşleşmesi gerekmektedir.

Bu çalışmamızda, ana çapa, çapa ve aday çapa olmak üzere 3 farklı tipte çapa düğüm kullanılmıştır. Bu çapa düğümler nitelik olarak tamamen birbiri ile aynıdır. Sadece farklı şekilde atamalar yapılabilir ve bu atamalar, adımlar içinde değişebilir. Ana çapa düğüm başlangıç durumunda ilk olarak başlanan ve uzayları birbirine bağlayan düğümdür. Aday çapa, ana çapa düğümü atandıktan sonra ana çapa düğümü etrafındaki düğümler arasında farklı iki uzayda bulunan ve birbirine her öznitelik yönünden bakıldığında eşik değerini geçen benzer düğüme verilen isimdir. Aday

düğüm işlem anından sonraki ilk anda eşik değerini geçiyorsa birleştirilip iki uzayı birleştiren tek düğüm olarak artık çapa olarak isimlendirilir.

Ana çapa seçimi yapılırken kullanıcıların kendi beyanları ile öğrenilen birden fazla sosyal ağdaki hesapları birbirine bağlanmıştır. Bu bağlama ile 2 uzay arası 2 farklı düğüm aynı düğüm olarak değişmiş ve 2 farklı düğümün tüm özellikleri tek düğümde toplanmıştır. Bir önceki bölümde bahsedilen farklı sosyal ağlar için türetilmiş veya özgün özniteliklerin eşleşme tablosu kullanılarak aynı öznitelikler tekilleştirilmiştir.

Ana Çapa düğümün birinci derece komşu düğümlerinin verileri toplanır. Toplanan tüm komşular aday çapa düğüm olarak adlandırılır. Aşağıda verilecek yöntemlerden biri kullanılarak aday çapa düğümler, diğer uzaydaki aday çapa düğümler ile benzerlik eşleştirilmeye çalışılır. Bu çalışmada 5 farklı yöntem önerilmiş ve test edilmiştir.

1) 2li Benzerlik Yöntemi (B-1) - Sadece Düğüm Benzerlik Tabanlı Düğüm Eşleştirme Yöntemi

2) 2li Hizalama Yöntemi (B-2) - Sosyal Ağlarda Düğüm Hizalama Yöntemi

3) 2li Topoloji Yöntemi (B-3)- Sadece Topolojik İlişki Tabanlı Düğüm Eşleştirme Yöntemi

4) 2li Ağırlıklı Topoloji Yöntemi (B-4)- Ağırlıklı Topolojik İlişki Tabanlı Düğüm Eşleştirme Yöntemi

5) 2li Hibrit Yöntemi (B-Hibrit)– Benzerlik ve Topolojik Tabanlı Düğüm Eşleştirme Yöntemi

Önerilen her yöntemin testi ayrı ayrı yapılmış bu bölümün son kısmında ise yöntemler detaylı olarak karşılaştırılmıştır.

1. Her deneyin yapılabilmesi için en az 60 adet çapa olarak eşleşebilir gerçek düğüm olması gerekmektedir.

2. Her deneyin sonuçlanması için en az ikinci seviye arkadaşların toplanması gerekmektedir.

3. İki düğümün karşılaştırılabilmesi için en az 4 özniteliğin ortak olarak yer alması gerekmektedir.

Düğümler arası ilişkinin puan olarak hesaplandığı yöntemler için bir etkileşim puan belirleme süreci önerilmiştir. Etkileşim yoğunluk puanı sosyal ağ bazında yapılan testler sonucunda normalize edilmiş ve 1 ile 2 arasında sınırlandırılmış puandır. Sadece düğümün topolojik uzaklık hesaplarında diğer düğüme gitmek için geçtiği düğümler için ağırlıklandırılmış metotlarda kullanılır. Sosyal ağlar için her bağlantının aynı uzaklıkta olduğunu söylemek mümkün değildir. Bundan dolayı ağırlıklandırılmış yöntemlerde kullanıcıların ilişkilerinin yoğun olduğu kullanıcılara daha yakın olduğunu görülür. Bu bağlamda önerilen yöntem her sosyal ağ için her bağlantı ile ortalama ilişki sayısını çıkarmakta ve 2 düğüm arasındaki ilişkiyi bu sayıya göre orantılanmakta ve normalleştirilerek 1 ile 2 arasında puan vermektedir. Ortalamayı bozduğu düşünülen ve çok sık görüşüldüğü öngörülen az sayıda kişi için filtre uygulanmakta ve ortalama hesaplanırken çok sık görüşülen bu kişiler hesaplamaya alınmamaktadır. Sosyal ağların yıllar içinde değişen oranda kullanım şekillerinden dolayı her sosyal ağ için uygulanan ortalama hesaplama yönteminin belli aralıklarla tekrar çalıştırılıp tekrar hesaplama yapılması gerekmektedir.

Önerilen Yöntemler

Dokuz farklı sosyal ağda ondan fazla öznitelik kullanıldığı için birden fazla yöntemin önerilmesi zorunluluğu doğmuştur. Önerilen dokuz yöntemden biri basitleştirilmiş başlangıç noktası test yöntemidir. Bu yöntemde sadece başlangıç noktasındaki değişimin başarı oranına etkisi ölçülmüştür. İki sosyal ağ için 5 yöntem ve ikiden fazla sosyal ağın kullanılacağı durumlar için 3 tane olmak üzere 8 tane de sosyal ağlarda aynı kişiyi tespit etmek için

yöntemler önerilmiştir. İkili sosyal ağlar için önerilen beş yöntemin ilk dördü, beşinci yöntemin (B-Hibrit) hazırlayıcısı niteliğindedir. Dört yöntemin derlenmesinden oluşan yöntem beşinci yöntem olarak sunulmuştur. İkili sosyal ağlar için kullanılan 5 yöntem temel alınarak hazırlanmıştır.

2li Benzerlik Yöntemi (B-1)- Sadece Düğüm Benzerlik Tabanlı Düğüm Eşleştirme Yöntemi

Sosyal ağlarda aynı kullanıcıları bulmak için de düğümler arası benzerlik ilişkisine bakmak önemli bir kriterdir. Birçok düğüm benzerlik yöntemi mevcuttur. Bu önerilen yöntem düğümlerin özniteliklerine odaklandığı için vektörel benzerlik yaklaşımını temel almaktadır. Böylece birçok özniteliğin bağımsız olarak vektörel benzerliği değerlendirilebilir. Bu durum aşağıda formülüze edilmiştir.

$$f_{2lib}(x) = \sum_{Öznitelikler} b(x) * \frac{U(a,b)}{1-\frac{d_T}{s_d}} \quad (1)$$

$f_{2lib}(x)$ ikili benzerlik değerini gösterirken, $b(x)$ iki düğüm arası seçilen özellik için benzerlik oranını temsil etmektedir. $U(x,y)$ X ve Y düğümleri arasındaki uzaklığı ifade eder. tüm uzaydaki iki düğüm arası uzunlukların toplamını verirken, ise düğüm sayısını ifade eder.

Seçilen ve karşılaştırılacak kullanıcılar arasında öznitelik değerlerinden biri her iki kullanıcıda da tanımlı değilse o öznitelik hesaba katılmaz. Öznitelik bazında uygulanan yöntemler Tablo 3'da gösterilmiştir.

Tablo 3. Öznitelik Bazında Uygulanan Yöntemler

Öznitelik	Uygulanan Yöntem
Kullanıcı İsmi	N-Gram Benzerlik Yöntemi
Lokasyon	Öklid Lokasyon Benzerlik Yöntemi
Kullanılan Dil	SimRank
Aktif Saat Dilimi	Keşisim Yöntemi
İlgi Alanı Etiketleri	Word N-Gram Benzerlik Yöntemi
Kişisel Bilgi	TF-IDF

Bu çalışmada, öznitelik 1 puan ile puanlandırıldı. Bundan dolayı tüm öznitelik benzerlik yöntemleri sonuçları en düşük 0 en yüksek 1 olacak şekilde düzenlendi. Her düğüm kendisi dışındaki tüm aday düğümler ile anlatılan hesaplama yapılarak eşik değeri üstündeki düğümler çapa düğüm olarak eşleştirilerek düğüm çiftleri birleştirilmiştir.

2li Hizalama Yöntemi (B-2)- Sosyal Ağlarda Düğüm Hizalama Yöntemi

Düğümlerin iki boyutlu bir uzay üzerinde birbiri ile ilişkilerini gösteren grafa iki boyutlu düğüm uzayı diyoruz. Düğümler, uzaylar üzerinde özniteliklerinin birbirine benzer olmasına ve birbiri ile ilişkilerinin yakın olmasına göre dizilirler. Düğümlerin birbiri ile ilişkili olanların komşu olarak gösterilmesi, benzer olanların ise yakın olarak gösterilmesi beklenir. Kullanıcılar farklı sosyal ağlarda farklı şekillerde davrandıkları için birden fazla sosyal ağdaki tüm düğümlerin aynı düzlem üzerinde tamamen aynı noktalarda olması beklenmemektedir. İki uzay üzerinde çapa düğümler birbiri ile ilişkilendirdikten sonra çapa düğümün etrafında kalan diğer düğümlerin çapa düğümlere uygun şekilde yayılmasına düğüm hizalama denir. Bu işlemin uygulanma adımları Şekil 1'de bulunan kabakod'da verilmiştir.

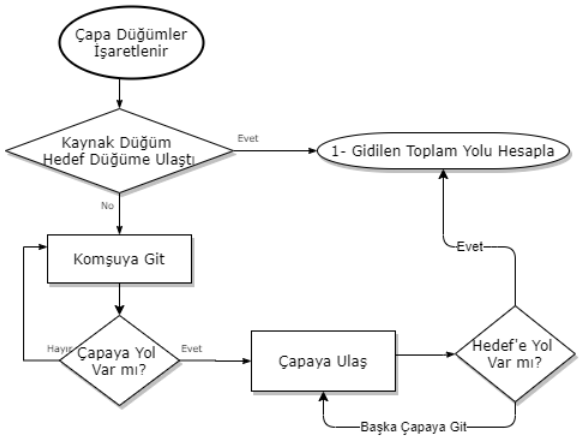
```
while(hizalama>esik degeri)
for each i in tum_dugumler.size
if (tum_dugumler[i].x > capa.x (and) tum_dugumler[i].y > capa.y)
tum_dugumler.[i].x++; tum_dugumler[i].y++
else if(tum_dugumler.[i].x > capa.x (and) tum_dugumler. [i].y < capa.y)
tum_dugumler.get[i].y--
else if(tum_dugumler[i].x < capa.x (and) tum_dugumler[i].y > capa.y)
tum_dugumler[i].x--;
end
end
```

Şekil 1. 2li Hizalama Sözde Kod Örneği

Örnek olarak, her iki uzayda da bulunan A noktaları bir çapa ile birbirine tutturulsun. A noktası ile B noktası arasında Uzay 1'de 2 eleman, Uzay 2 de ise 5 eleman olsun. Her iki uzaydaki B noktaları da çapa olarak tespit edildiği anda A ve B düğümleri arasında 2 eleman olan uzay 5 eleman olan uzaya benzemek için genişleyecektir. Bunun nedeni her iki uzaydaki B'ler birbiri ile ilişkilendirilerek Uzay 1'deki B, Uzay 2'deki B hizasına çekilecektir. Bu süreçte B'ye yakın olan tüm düğümlerin de orantılı olarak kayması beklenir.

2li Topoloji Yöntemi (B-3)- Sadece Topolojik İlişki Tabanlı Düğüm Eşleştirme Yöntemi

Tek başına düğüm hizalama aynı kullanıcılara ait hesapları bulmak için doğru bir yöntem olarak kullanılmamaktadır. Bunun nedeni, uzayların graf şeklinde dağıtıldıktan sonra tek çapa düğüm ile bağlandığı durumlarda aynı yere hizalanan 2 düğümün farklı kişiler olabileceği gerçeğidir. Sosyal ağlarda benzerlik bulunurken sadece düğümün benzerliğine veya sadece topolojiye bakmak eksik bir yöntem olarak kabul edilebilir. Düğüm benzerliği ile birlikte ağın topolojisi ve sunduğu ilişki grafiğine bakılarak da yapılan bir düğüm benzerlik yöntemi daha doğru sonuç verecektir. Bu yöntemin akış diagramı Şekil 2'te sunulmuştur.



Şekil 2. 2'li Ağırlıklandırılmış Topoloji Yöntemi Akış Diagramı

Böylece ilişkisi çok kuvvetli düğüm üzerinden geçerken yakınlık puanı yüksek iken, ilişkisi çok zayıf olan düğüm üzerinden geçtiğinde daha düşük bir puan alacaktır. Böylece puanlandırma işlemleri iki düğüm arasındaki etkileşim yoğunluğuna göre hesaplanmıştır.

2li Hibrit Yöntemi (B-Hibrit)- Hibrit Düğüm Eşleştirme Yöntemi

Bu yöntem '2li Benzerlik' ve '2li Topoloji' yöntemlerinin birleştirilmiş bir versiyonudur. '2li Benzerlik' ve '2li Topoloji' yöntemlerine ek olarak bir normalleştirme fonksiyonu ile uygulanan yöntemlerden gelen sonuçların doğru şekilde birleşmesi sağlanmıştır. Bu yöntem, 2li Benzerlik' yöntemini sayesinde değerlendirilen ve eşik değerinin üstünde yakınlık sağlayan düğümlerin bir listeye eklenmesi ile başlanır. Listedeki düğümler '2li Topoloji' yöntemine göre uzaklıkları tespit edilip puanlarına eklenir. Normalleştirme yapılırken hem benzerlik puanı hem de topolojik puanı birleştirerek yeni puan ortaya çıkartılır. Benzerlikten ve topolojik puandan gelen iki değerin ortalaması alınarak eşik değeri geçmesi beklenir. Bu normalleştirme ile eğer benzerlik değeri eşik puanın yarısını geçememişse veya topolojik yakınlık puanı yarım puanı geçmemişse ne kadar benzer veya yakın olursa olsun eşleşme sağlanmayacaktır. Bu ölçüt aşağıdaki şekilde formülize edilir.

$$f(x) = \sum_{ÇAD_S} \sum_{AÇD_S} Z(f_{2lib}(x,y), d(x,y)) \qquad (2)$$

$AÇD_S$ seçilen düğümün aday çapa düğümlerin listesini, $d(x,y)$ ise tüm çapa adayı düğümleri listesini, ise düğümler arası uzunluğu gösterir.

Bu yöntem bazı düğümleri tespit edemeyebilirken kesinlik açısından güvenilir bir yol izlemektedir. Her uzay değişiminde topolojik hesaplama tekrar yapılır ve yapıldıktan sonra benzer olan ama topolojik puanı yeterli olmayan düğümler yeni oluşan çapa düğümlerin üzerinden geçebilir ve aday düğümler ile aralarındaki puan değişebilir. Her değişimde ağda oluşan yeni bağlantılardan dolayı topolojik puan sabit kalmayacak ve yükselecektir. İki düğüm arası benzerlik puanı ise hep aynı

kalarak güven değerini oluşturmaktadır. Bu yöntemde topolojik puan sürekli değişeceği için bir son çapa ekleme sınırı koyulması gerekmektedir. Yapılan testlerde fazla iterasyonun hatalı değerler ortaya çıkarmadığı ama zaman kaybına yol açtığı görülmüştür. Bundan dolayı iterasyon sayısı düğüm sayısının log(n) karşılığı olarak belirlenmiştir. Üst limit log(n) olsa bile eğer son iterasyonda değişiklik olmamışsa ağ değişmeyeceği için iterasyon sonlanır ve sonuç kesinleşir. Önerilen yöntem bu çalışma dışındaki yöntemler ile karşılaştırılacaktır.

Deneysel Sonuçlar

Çalışmanın bu bölümünde önerilen bir ilkel ve sekiz gelişmiş yöntem teste tabi tutulmuştur. Çalışmada önerilen yöntemler ile aynı işi yapan muadil bir algoritma olmadığı için tüm yönleri ile başka bir yöntem ile test edilememiştir. Sadece tek uzayda topolojik hizalama ve geçişlilik ile ilişki bulma metotları önceki çalışmalarla karşılaştırılmıştır. Diğer tüm önerilen yöntemler testi, test kümelerinin manuel olarak gönüllüler tarafından işaretlenmesi ile oluşturulan kontrol grubuna göre test edilmiştir.

Test Yöntemleri

Önerilen yöntemlerin çoğunda karşılaştırma için muadil bir yöntem bulunmadığından biri hariç tüm yöntemlerde başarı oranları bağlantı eksiltmeyle teste edilmiştir.

İlk olarak topolojik hizalama ve geçişlilik yöntemlerinin testi yapılmıştır. Topolojik hizalama hizalanmış iki uzayı birbiri ile karşılaştırarak, geçişlilik ise aynı geçiş yollarının bulunup bulunmadığının tespiti ile yapılır. Geçişlilik yol tespitinde mevcut önerilen yöntemlerde olmadığı için çapalama yapılmaz. İlişkiler üzerinden geçişler takip edilir. Bundan dolayı geçişliliğin testi yapılmış olsa da bu çalışmanın konusu olan uzay birleştirmenin testinin yapıldığı söylenemez. Topolojik hizalama ise önceki çalışmalarda özetle ilişkilerin ağırlığı ve özellik vektörüne göre düğümlerin uzay üzerine hizalanarak dağıtımı olarak ifade edilebilir. Çalışmamız topolojik hizalama ve puanlamanın temelini oluşturduğu için bu testteki başarı oranları çalışmamız için önemlidir.

Karşılaştırma yapabilmek için çoklu sosyal ağlarda düğüm hizalama yapan yöntemlerden 3 tanesi seçilerek önerdiğimiz hizalama yöntemi ile test yapılmıştır. Karşılaştırma yapılan diğer yöntemler genelde ikili sosyal ağ kullandığı için bu çalışmada yer alan '2li Hizalama' yöntemi test için seçilmiştir. Test kümesi olarak da Instagram ve Twitter verileri kullanılmıştır. Hyrdra [18] yönteminde kullanılan görüntü işleme ve yüz tanıma konuları bu çalışmada yapılan testte kullanılmamıştır. Karşılaştırma için kullanılan algoritmalar ile yapılan testlerin sonuçları Tablo 4'de sunulmuştur.

Tablo 4. Topolojik Hizalama Yöntemlerinin Karşılaştırmalı Başarı Oranları

Yöntem İsmi	A	P	R	F
2li Hizalama	0,77	0,92	0,77	0,84
Liu – Cheung [14]	0,60	0,85	0,62	0,71
Hyrdra [18]	0,60	0,71	0,67	0,69
Wu-Chien [25]	0,60	0,78	0,58	0,67

Tablo 4'te görüldüğü gibi en iyi sonuçlar önerdiğimiz yöntem ile elde edilirken, ikinci en iyi sonuçlar Liu – Cheung yaklaşımı ile elde edilmiştir. En kötü sonuçlar ise P değeri hariç Wu-Chein'in yöntemi ile elde edilmiştir.

Çalışmamızda ilk olarak bu makalenin yazarlarından biri olan ve ilk analizde seçilen tüm sosyal ağlarda hesabı bulunan 'Ahmet Müngen' kullanıcısının hesabı temel alınarak veri toplanmaya başlanmıştır. Veri toplama işlemi daha önceki bölümlerde detaylı bir şekilde bahsedilmiştir. Kriterlere göre ilk kullanıcının ve o kullanıcının bağlantılarının

verileri çekilerek veri toplama işlemine başlanmıştır.

Bağlantı Eksiltme Test Yöntemi

Bağlantı eksiltme testi, graf üzerinde bilinen tüm çapalar bulunup bağlandıktan sonra çapaların bir kısmının rastgele silinmesi sonrası sistemin aynı çapayı bulmaya çalışmasıdır. Bağlantı eksiltme testlerinde düğümlerin hiçbir özelliği değiştirilmez veya silinmez. Bunun yerine ilgili çapa bağlantısı kaldırılır ve sistemden çapa bulması istenir. Bu test en güvenilir test olmakla birlikte hangi düğümlerin kaldırılacağı testin başarısı ile birebir orantılıdır. Bu testte düğüm kaldırma işlemi aynı düğümün varsa ikinci derece ve üstü uzaklıktaki düğümlerinin kaldırılması ile başlar. Böylece düğümü birebir etkileyen çapa kaldırılmadığı için düğüm bulma işlemi daha başarılı olur. Eksiltilen çapa sayısı arttıkça ise ikinci ve üstü uzaklıktaki çapalar kalmayacak ve birinci dereceden yakın çapalar eksiltilmeye başlanacaktır. Birinci dereceden eksiltilen her çapa sonucu daha çok olumsuz etkileyecektir. Karşılaştırmaların daha anlaşılır olması için bazı testler birden fazla tabloda geçmektedir. Tablo 5'te B ve C tipi yöntemler için yapılan bağlantı eksiltme testlerinin sonuçları gösterilmektedir.

Tablo 5. B ve C-Tipi Metotlar Bağlantı Eksiltme Testleri

Seçilen Sosyal Ağlar	Yöntem	Çıkarılan Çapa	A	P	R	F
Instagram Linkedin	B-Hibrit	1 Adet	0,98	0,99	0,99	0,99
Instagram Linkedin	B-Hibrit	10%	0,95	0,97	0,96	0,97
Instagram Linkedin	B-Hibrit	20%	0,94	0,96	0,96	0,96
Instagram Linkedin	B-Hibrit	30%	0,92	0,95	0,95	0,95
Instagram Linkedin	B-Hibrit	40%	0,91	0,95	0,94	0,95

Tablo 5'de görüldüğü gibi değerler belirli sayıda düğümün çıkarılmasıyla elde edildiği için genel olarak başarı oranları yüksektir. Ayrıca bu test yapılırken tüm düğümler doğru çapalanmış ve tamamen ideal bir sistem üzerinden çıkartılarak yapıldığı için diğer testlerden daha başarılı olacağı açıktır. Ayrıca çıkarılan çapa seçimindeki kurallar da sonucun yüksek çıkmasını desteklemiştir. Tek çapa çıkarıldığında çok büyük oranda doğru bulurken çıkarılan çapa sayısı arttıkça başarı oranın düştüğü görülmüştür. Buna karşın %40'ı çıkarılsa bile başarı oranının yüksek olduğu görülmüştür.

Çapa Düğüm Bulma Testleri

Bu testte bazı çapa kullanıcıları tamamen her iki uzaydan kaldırılmıştır. Önerilen yöntem çalıştırılarak çıkartılan çapa düğüm dışında aynı sonuca ulaşıp ulaşılmadığı tespit edilmiştir. Bu yöntem uygulanırken tespit edilen tüm çapaların en fazla yarısı çapalanır diğer yarısı serbest bırakılır. Bu test için sosyal ağlar kombinasyonlu olarak seçilip veri setine dönüştürülerek kısmi veri setleri oluşturulmuştur. Önerilen ikili kombinasyonlu veri setleri için 'B-Hibrit' yöntemi kullanılarak yöntemlerin sosyal ağlar bazında başarı oranları tespit edilmeye çalışılmıştır. Aynı zamanda seçilen veri tabanının sayısının genel başarıya oranı da hesaplanmıştır.

Bu testler yapılırken farklı veri tabanları birleştirilerek veri setlerine dönüştürülmüş ve veri seti üzerinde bir kullanıcı çapalanarak çapa başlangıç durumu belirlenmiştir. Bu adımdan sonra çapa bulma yöntemi çalıştırılıp test sonuçları kaydedilmiştir. Tablo 6'da yapılan testlerden bazı sonuçlar paylaşılmıştır.

Tablo 6. Örnek Test Sonuçları

Sosyal Ağ Kombinasyonu	A	P	R	F
Instagram Linkedin	0,900	0,957	0,938	0,947
Linkedin Twitter	0,900	0,967	0,927	0,947

About.me ile Çapa Kaldırma

About.me Ekim 2009'da kurulan kişisel profil listeleme hizmetidir. Site, kayıtlı kullanıcılara birden fazla çevrimiçi sosyal ağ bilgisini ve Facebook, Flickr, Google+, Pinterest, LinkedIn, Twitter, Tumblr ve YouTube gibi popüler sosyal ağ web sitelerindeki kendi profillerini bağlamak için basit bir platform sunmaktadır. Özelleştirilebilir arka plan görüntüsü ve kısaltılmış biyografiye sahip tek sayfalık kullanıcı profilleri ile de hizmet sunmaktadır. About.me profil görüntüleme ve profil incelemeleri alanında daha önce de farklı akademik çalışmalarda [26][27] kullanılmıştır.

Tablo 7. About.me Profil Bulma Testleri

Sosyal Ağ	Düğüm Tipi	A	P	R	F
Twitter - Instagram	Tek Çıkarılan Ana Çapa Düğüm Bulma	0, 95	0, 98	0, 97	0, 97
Twitter – LinkedIn	Tek Çıkarılan Ana Çapa Düğüm Bulma	0, 94	0, 99	0, 95	0, 97
Twitter-Flickr	Tek Çıkarılan Ana Çapa Düğüm Bulma	0, 90	0, 96	0, 94	0, 95
İnstagram - LinkedIn	Tek Çıkarılan Ana Çapa Düğüm Bulma	0, 96	0, 97	0, 99	0, 98
Twitter – Instagram	Serbest Çevre Düğüm Bulma	0, 79	0, 84	0, 90	0, 87
Twitter – LinkedIn	Serbest Çevre Düğüm Bulma	0, 73	0, 84	0, 83	0, 84
Twitter-Flickr	Serbest Çevre Düğüm Bulma	0, 77	0, 81	0, 91	0, 86
İnstagram - LinkedIn	Serbest Çevre Düğüm Bulma	0, 80	0, 82	0, 87	0, 85

Toplanan profiller arasında açık profili olan 2.879 kullanıcı teste tabi tutulmuştur. Daha önceki bölümlerde bahsedilen ve 'Ahmet Müngen'in profili çevresinde kurgulanan test kümesi buraya uygun olmadığı için yine 2.879 kullanıcının her biri ve çevresi için tekrar veri toplama işlemi yapılmıştır. Böylece kullanıcının Twitter arkadaşları ile Facebook arkadaşları eşleştirilmeye çalışılmıştır. Bu testin sonucu manuel kontrol ile tespit edilmiştir. Manuel kontrol sonrası onaylanan yeni çapa düğümlerden sonra about.me profili olan açık profilli kullanıcıların çapaları kaldırılmış ve sistem teste tabi tutulmuştur. Bu test ile önerilen yöntemin sadece Türkiye'den Türkçe konuşan kullanıcılarda değil tüm dünyadaki kullanıcılarda başarı ile çalıştığını ortaya koymuştur.

Tablo 7 'Serbest Çevre Düğüm Bulma' işlemi ile çapa düğüm seçildikten sonra kullanıcının iki farklı sosyal ağda arkadaşları arasında düğüm bulma uygulamasının sonuçlarını göstermektedir. 'Tek Çıkarılan Ana Çapa Düğüm Bulma' işlemi ise çevre düğümler bulunup çapalandıktan sonra ana düğümün çıkarılıp test yapılması işlemidir. 'Tek Çıkarılan Ana Çapa Düğüm Bulma' işlemi sırasında çıkartılan ana düğümün etrafında referans olabilecek çok sayıda çapa düğüm olduğu için bu profilin bulunma oranı çok yüksektir.

Profil bulma oranlarının sosyal ağ sayısı arttıkça arttığı görülmektedir. Ayrıca LinkedIn sosyal ağının içerdiği testlerde diğer sosyal ağlara göre kısmen daha iyi olduğu görülmektedir. Bunun nedeni LinkedIn profillerindeki bilgilerin diğer sosyal ağlara göre daha özenli / güncel ve güvenilir olmasıdır.

Yöntemlerin Kendi Aralarında Değerlendirme Sonuçları

Önerilen tüm yöntemler kendi aralarında da başarı değerlendirmesine tabi tutulmuşlardır. Bu testlerin amacı yöntemleri kendi içinde de değerlendirmek ve eklenen yeni özelliklerin yöntemlerin başarı değerleri üzerinde ne kadar etkili olduğunu tespit etmektedir. Aday Çapa yöntemi diğer yöntemlerden çok farklı olduğu için bu teste tabi tutulmamıştır. Testler Instagram – Twitter veri kümesi içerisinde yapılmıştır. Test sonuçlarının değerlendirmesinin objektif olması için her test için 100 adet düğümün var olduğu Instagram – Twitter verilerinden oluşan bir veri kümesi hazırlanmıştır. Yapılan testler ile ilgili sonuçlar Tablo 8'da verilmiştir.

İkili Sosyal Ağ Değerlendirme Sonuçları

Tablo 9 dokuz adet sosyal ağın B-Hibrit yöntemi ile ikili kombinasyonlarının sistem üzerinden yapılan testlerin sonuçlarını içermektedir. Burada ikili kombinasyonlardan otuz altı farklı seçenek için test yapılmış ve test sonuçları detaylı şekilde verilmiştir.

Tablo 8. Önerilen Yöntemlerin Kendi Aralarında Karşılaştırma Değerleri

Önerilen Yöntem / Değerlendirme Yöntemi	A	P	R	F
B- Benzerlik	0,72	0,94	0,67	0,78
B- Hizalama	0,77	0,92	0,77	0,84
B- Topoloji	0,78	0,86	0,84	0,85
B- Ağırlıklı Topoloji	0,81	0,90	0,86	0,88
B Hibrit	0,86	0,93	0,91	0,92

Tablo 9. İkili Sosyal Ağlar için B-Hibrit Deney Sonuçları

Sosyal Ağ Kombinasyonu	A	P	R	F
Instagram LinkedIn	0,90	0,95	0,93	0,94
LinkedIn Meetup	0,87	0,94	0,91	0,93
Instagram Twitter	0,86	0,93	0,91	0,92
Flickr LinkedIn	0,86	0,91	0,91	0,91
Instagram Meetup	0,85	0,92	0,91	0,91
LinkedIn Pinterest	0,83	0,89	0,90	0,90
Foursquare Twitter	0,82	0,91	0,87	0,89

Tablo 9'de 36 sonuçtan en başarılı 10 sonuç paylaşılmıştır. Bu tabloda Linkedin ve Instagram sosyal ağlarının düğüm tespitinde en başarılı ağlar olduğu bunu nedeninin de verilerinin diğer sosyal ağlara göre daha ayırt edici ve daha doğru olmasıdır. Twitter ise algoritmaya en çok başarı sağlayan üçüncü sosyal ağ olmuştur.

Sonuç

Bu çalışmada düğüm benzerlik ve hizalama yöntemleri ile ilgili özgün, kapsayıcı ve başarılı bir yöntem dizisi önerilmiştir. Önerilen bu bir dizi yöntem ile farklı öznitelik ve farklı özellikleri olan birçok sosyal ağ üzerinde düğüm benzerliği ve hizalama çalışılabilecektir. Ayrıca çalışmanın mevcut yöntemlere sunduğu farklı yaklaşımlar ile de düğüm hizalama konusunda özgün bir hesaplama yöntemi önerilmiştir. Çalışma sonucunda çoklu sosyal ağlardaki verilerin tek bir uzayda birleştirilebildiği gösterilmiştir. Böylece sosyal ağlarda çalışan neredeyse tüm algoritma ve yöntemlerin başarısını arttırabilecek çoklu sosyal ağın kullanımının önü açılmıştır.

Farklı amaçlar için kullanılan ve farklı kategorilerde değerlendirilen dokuz farklı sosyal ağdan toplanan verilerde öznitelik analizi ve karşılaştırması ile hangi özniteliklerin eşleştirilebileceği ve sosyal ağların diğer sosyal ağlar ile uyumu gibi analizler ve sonuçlar sunulmuştur. Oluşturulan veri seti çok çeşitli sosyal ağları kapsadığı için daha sonraki araştırmacıların da kullanabilecekleri ortak bir veri kümesi oluşturulması sağlanmıştır.

Sonraki çalışmalarda bu çalışmada ele alınmayan diğer özniteliklerin de kullanımı ile daha gelişmiş bir model önerilebilir. Yıllar içinde insanların kendilerini tanımlama süreci, ilgi alanları ve sosyal ağı kullanış biçimi değişeceği için sosyal ağların evrimleşmesi ve kullanıcıların hareketlerinin zaman içindeki değişimi de ele alınarak benzerlik ve hizalama algoritma önermeleri yapılabilir.

Teşekkür

Bu çalışma TÜBİTAK tarafından 119E309 numaralı proje kapsamında desteklenmiştir.

Kaynaklar

[1] J. Du, C. Jiang, K. C. Chen, Y. Ren, and H. V. Poor, "Community-structured evolutionary game for privacy protection in social networks," IEEE Trans. Inf. Forensics Secur., vol. 13, no. 3, pp. 574–589, Mar. 2018, doi: 10.1109/TIFS.2017.2758756.

[2] D. Koutra, H. Tong, and D. Lubensky, "Big-Align: Fast bipartite graph alignment," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2013, doi: 10.1109/ICDM.2013.152.

[3] L. Adamic and E. Adar, "How to search a social network," Soc. Networks, 2005, doi: 10.1016/j.socnet.2005.01.007.

[4] J. Jaccard, Interaction Effects in Logistic Regression. 2011.

[5] B. Aleman-Meza et al., "Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection," in Proceedings of the 15th International Conference on World Wide Web, 2006, doi: 10.1145/1135777.1135838.

[6] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," SIAM Rev., 2004, doi: 10.1137/S0036144502415960.

[7] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms , Second Edition. 2001.

[8] B. Cai, H. Wang, H. Zheng, and H. Wang, "An improved random walk based clustering algorithm for community detection in complex networks," in Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2011, pp. 2162–2167, doi: 10.1109/ICSMC.2011.6083997.

[9] F. Fouss, A. Pirotte, J. M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," IEEE Trans. Knowl. Data Eng., 2007, doi: 10.1109/TKDE.2007.46.

[10] P. Symeonidis and E. Tiakas, "Transitive node similarity: Predicting and recommending links in signed social networks," World Wide Web, vol. 17, no. 4, pp. 743–776, Jun. 2014, doi: 10.1007/s11280-013-0228-2.

[11] G. Jeh and J. Widom, "SimRank: A measure of structural-context similarity," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002.

[12] A. Clauset, C. Moore, and M. E. J. Newman, "Hierarchical structure and the prediction of missing links in networks," Nature, vol. 453, no. 7191, pp. 98–101, 2008, doi: 10.1038/nature06830.

[13] K. Musiał and P. Kazienko, "Social networks on the Internet," World Wide Web, 2013, doi: 10.1007/s11280-011-0155-z.

[14] L. Liu, W. K. Cheung, X. Li, and L. Liao, "Aligning users across social networks using network embedding," in IJCAI International Joint Conference on Artificial Intelligence, 2016.

[15] Y. Dong et al., "Link prediction and recommendation across heterogeneous social networks," in Proceedings - IEEE International Conference on Data Mining, ICDM, 2012, doi: 10.1109/ICDM.2012.140.

[16] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via Cross-Domain Triadic Factorization," in WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web, 2013, doi: 10.1145/2488388.2488441.

[17] J. Liu, F. Zhang, X. Song, Y.-I. Song, C.-Y. Lin, and H.-W. Hon, "What's in a Name? An Unsupervised Approach to Link Users across Communities *," Feb. 2013.

[18] S. Liu, S. Wang, F. Zhu, J. Zhang, and R. Krishnan, "HYDRA: large-scale social identity linkage via heterogeneous behavior modeling," in Proceedings of the 2014 ACM SIGMOD international conference on Management of data - SIGMOD '14, 2014, doi: 10.1145/2588555.2588559.

[19] J. Zhang and P. S. Yu, "Integrated anchor and social link predictions across social networks," in IJCAI International Joint Conference on Artificial Intelligence, 2015.

[20] S. Tan, Z. Guan, D. Cai, X. Qin, J. Bu, and C. Chen, "Mapping Users across Networks by Manifold Alignment on Hypergraph," 28th AAAI Conf. Artif. Intell., 2014.

[21] A. A. Mungen, A. Geckil, and M. Kaya, "Social Network Attribute Analysis Method for Node Alignment Process," Jan. 2020, pp. 1–4, doi: 10.1109/ubmyk48245.2019.8965616.

[22] S. Zhang and H. Tong, "Attributed Network Alignment: Problem Definitions and Fast Solutions," IEEE Trans. Knowl. Data Eng., 2018, doi: 10.1109/TKDE.2018.2866440.

[23] H. T. Trung, T. Van Vinh, N. T. Tam, H. Yin, M. Weidlich, and N. Q. Viet Hung, "Adaptive network alignment with unsupervised and multi-order convolutional networks," 2020, doi: 10.1109/ICDE48307.2020.00015.

[24] Q. Zhan, J. Zhang, and P. S. Yu, "Integrated anchor and social link predictions across multiple social networks," Knowl. Inf. Syst., 2019, doi: 10.1007/s10115-018-1210-1..

[25] S. H. Wu, H. H. Chien, K. H. Lin, and P. S. Yu, "Learning the consistent behavior of common users for target node prediction across social networks," in 31st International Conference on Machine Learning, ICML 2014, 2014.

[26] M. Wang, Q. Tan, X. Wang, and J. Shi, "De-anonymizing social networks user via profile similarity," in Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018, Jul. 2018, pp. 889–895, doi: 10.1109/DSC.2018.00142.

[27] M. M. Rahman, "Intellectual knowledge extraction from online social data," in 2012 International Conference on Informatics, Electronics and Vision, ICIEV 2012, 2012, pp. 205–210, doi: 10.1109/ICIEV.2012.6317392.