Anatolian Journal of Computer Sciences

Volume:5 No:2 2020 © Anatolian Science

pp:114-125 ISSN:2548-1304

Graf Benzerliği İle Metin Kiyaslama

Harun DARBAŞ[1*], *Ali KARCI*[2]

[*1]*Adıyaman Üni. TBMYO Bilgisayar Programcılığı, Adıyaman, Türkiye (hdarbas@adiyaman.edu.tr)*

[2] *İnönü Üni. Mühendislik Fak. Bilgisayar Mühendisliği, Malatya, Türkiye (ali.karci@inonu.edu.tr)*

Received Date : May 28, 2020.

Acceptance Date : Oct. 20, 2020.

Published Date : Dec. 1, 2020

Graf benzerliği NP-zor olan bir problemdir ve metin benzerliği problemini çözmekte aynı şekilde yaklaşım gerektiren bir problemdir. Günümüzde çok farklı alanlarda graf benzerliği kullanılmaktadır. Bu konu yaklaşım yöntemlerle çözülmeye çalışıldığından graf benzerliği ölçümleri de birbirinden farklılık göstermektedir. Yeni bir graf benzerliği ölçümü ortaya konularak daha önce kullanılan alanlardan farklı olarak metin benzerliğinin ölçülmesinde kullanımı amaçlanmaktadır.

Bu çalışmada, daha önce düğüm bezerliği hesabıyla ve düğümlerin kıyaslanmasıyla ölçülen graf benzerliğinin, grafların yapısal özelliklerinin kıyaslanmasıyla ölçülmesi amaçlanmaktadır. Bu benzerlik durumu metin benzerliği için kullanılmıştır ve çalışmanın sonuçları bu makalede değerlendirilmiştir.

Keywords : *Graf Benzerliği, Metin Madenciliği, Metin Benzerliği.*

1. GİRİŞ

Günümüzde çok farklı alanlarda graf benzerliği kullanılmaktadır. Graf benzerliği ölçümleri de birbirinden farklılık göstermektedir. Yeni bir graf benzerliği ölçümü ortaya konularak daha önce kullanılan alanlardan farklı olarak metin benzerliğinin ölçülmesinde kullanımı amaçlanmaktadır.

Graf benzerliği üzerinde yapılan çalışmalarda kesin çözüm veren bir algoritma geliştirilmiş değildir ve bundan dolayı farklı yöntemlerle bu problem üzerinde çalışmalar yapılmıştır. Bunların bir kısmı aşağıda listelenmiştir.

Graf benzerliğini, Vrotsou (2009) çalışmasında, verilerin kullanıcı merkezli olarak araştırılmasını ve o kullanıcı için önemli verilerin tanımlanmasını kolaylaştırmak amacıyla sezgisel bir görsel arayüzle birleştirip web araması için geliştirilen tekniklerin adaptasyonuna dayanan etkileşimli bir görsel veri madenciliği yaklaşımında kullanmaktadır. Benzerlik ölçümü için düğüm benzerliklerinden faydalanılmış olup bir graftaki j düğümünün yetki skoru, j düğümü ile o düğümün düğüm yetkisi arasındaki benzerlik skoru olarak görülebileceği belirtilmiştir.

Koutra (2011), graf benzerliği ve alt graf eşleştirme konulu çalışmasında, sezgisel olarak, düğüm benzeşmelerini bilindiğinde her iki graftada aynı düğüm, komşuları benzerse ve kenar ağırlıkları açısından, komşularıyla bağlantısının benzer olacağını belirtmiştir. Koutra (2013), bu sorunu çözmenin bariz yolu olarak kenarlarının üst üste gelmesini ölçmek şeklinde değerlendirmeye çalışmıştır. Çalışmasının devamında, önerilen benzerlik algoritmasının arkasındaki ana fikrin, verilen graflarda düğüm afinitelerini karşılaştırmak olduğuna değinerek, sıraladığı 3 adım ise şöyledir; Adım 1: her bir graf için çiftli düğüm afinite skorlarının n x n matrisi hesaplanır. Adım 2: kök öklid mesafesi kullanılır

(ROOTED). Adım 3: yorumlanabilirlik için, $sim = \frac{1}{1+d}$ formülüyle d mesafesi, benzerlik ölçüsüne (sim) dönüştürülür.

Zhao (2012) graf benzerliğine katılma problemini, yollara dayalı yeni bir q-gram kavramı getirerek çözmeye çalışmıştır. Ağaç bazlı q gram sayısı, sayım filtreleme koşulu geliştirilmiştir.

HMM (Hidden Markov Models) tabanlı Tarihsel belgelerde el yazısı tanıma çalışmasında Fischer (2010), metin görüntülerinin yapısal graf tabanlı temsilini kullanarak, karakter prototipine göre farklılık katma yoluyla bir dizi graf benzerlik özelliği çıkarmıştır.

Öğrencilerin programlama ödevlerinin otomatik işaretlemesine yeni bir yaklaşım sunan Naudé (2010), graf benzerliği kullanarak işaretlenmemiş öğrenci sunumları ile işaretli çözümler arasındaki yapısal benzerliği nicelendirip not ataması yapmıştır. Graf benzerliği ölçümünde düğüm ve komşularının benzerliğinden faydalanmıştır.

Moleküler benzerlik ölçümü için graf benzerliğinden faydalanan Skvortsova (1998), rastgele etiketlenmiş graflar için tanımlanan simetrik benzerlik kümelerinin analitik bir açıklamasını vermiştir.

Opcode graflara dayalı olarak çalıştırılabilir dosyaların benzerliğini hesaplamak için bir yöntem düşünen Runwal (2012), graf tabanlı benzerlik saptaması için HMM kullanmıştır.

Bir graftaki bir düğümün diğer grafın bir dizi düğümü ile ilişkilendirilebilmesi için çok değerlikli eşlemelere dayalı bir benzerlik ölçüsü sunan Sorlin (2005), graf benzerliği için Reaktif Tabu Arama kullanmıştır.

Literatür incelemesinde graf benzerliğinin bunların dışında, anomali tespiti, metamorfik algılama, görüntü tanıma, insan sınıflandırma (beyin bağlantı grafına göre yüksek ve düşük yaratıcılık), web arama, virüs algılama, güven tabanlı tavsiye, öğrenci ödevlerinin değerlendirilmesi, yüz tanıma ve kötü amaçlı yazılım tespitinde de kullanıldığı görülmüştür.

Bu çalışmada, daha önce düğüm bezerliği hesabıyla ve düğümlerin kıyaslanmasıyla ölçülen graf benzerliğinin, grafların yapısal özelliklerinin kıyaslanmasıyla ölçülmesi amaçlanmaktadır.

Metinlerde; cümleler düğümleri, kenarlar ise cümleler arası ortak kelime sayısını gösterecek şekilde yönsüz ve ağırlıklı graflarla modellenmektedir. Bu grafların bazı yapısal özellikleri çıkarılmakta ve nihayetinde bu özellikler kullanılarak metin benzerlik oranı hesaplanmaktadır.

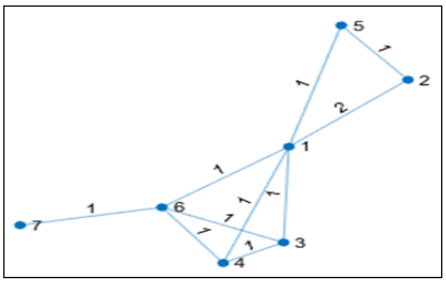
2. GRAF VERİ MODELİ İLE METİN MODELLEME

Bir G=(V,E) grafı, V ile gösterilen düğümlerden (node veya vertex) ve E ile gösterilen kenarlardan (edge) oluşur. Soyut nesneler, köşeler ve kenarlardan oluşan bir koleksiyondur (Spizzirri, L. 2011).

Köşeler veya düğümler olarak adlandırılan öğelerin, V(G) gibi boş olmayan bir sonlu kümesi ile kenarlar olarak adlandırılan V(G)' nin farklı öğelerinin sıralanmamış farklı öğe çiftlerinin sonlu bir kümesi olan E(G)' yi içeren basit bir G grafını veri modelleme için kullanılmaktadır. V(G) ye düğümler kümesi ve E(G) ye de G nin kenarlar kümesi denir. Bir {u, v} kenarı u ile v düğümlerini birleştirir ve genellikle uv olarak kısaltılır (Robin JW, 1996).

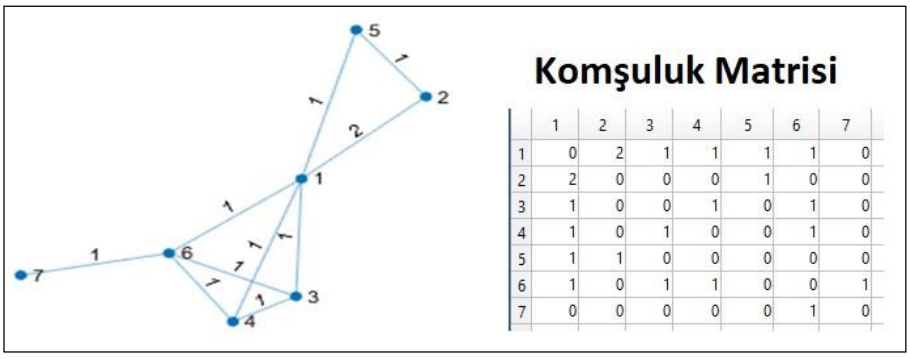
G' de bir (u, v) yolu varsa u ve v köşelerinin birbirine bağlı olduğu söylenir. Bağlantı, V düğüm kümesi üzerinde bir denklik ilişkisidir. V' nin boş olmayan V_1, V_2, ... , V_n alt kümelerinin içinde bir parça olarak, sadece ve sadece u ve v aynı alt kümeye aitse bu iki düğüm u ve v, bağlıdır denir. $G[V_1]$, $G[V_2]$, ... , $G[V_n]$ alt grafları G' nin bileşenleridir. G' nin tam olarak sadece bir bileşeni varsa, G bağlı graftır; aksi takdirde değildir (John AB, 1982).

Eğer G grafının kenarlarını gösteren E kümesinin tüm elemenalarına ait birer ağırlık değeri varsa bu tür graflara ağırlıklı graflar denir. Aynı şekilde bu kenarların yönlerinin olmaması durumundada Şekil 1' de ki gibi yönsüz ağırlıklı graf türüne ait olurlar.



Şekil 1. Yönsüz ağırlıklı graf

G grafının u ile v düğümlerini birleştiren bir uv kenarı varsa, bu düğümler komşudur. Benzer şekilde, eğer iki kenarın birer düğümleri ortak ise, bu kenarlar da komşudur. Bir grafa ait komşu düğümlerin *nxn* boyutlu bir matriste sunulmuş haline komşuluk matrisi denir. Şekil 2 de, verilen grafın komşuluk matrisi oluşturulmuştur.



	1	2	3	4	5	6	7
1	0	2	1	1	1	1	0
2	2	0	0	0	1	0	0
3	1	0	0	1	0	1	0
4	1	0	1	0	0	1	0
5	1	1	0	0	0	0	0
6	1	0	1	1	0	0	1
7	0	0	0	0	0	1	0

Şekil 2. Verilen grafa ait komşuluk matrisi

Metinlerin temsil edilmesi için veri yapısı olarak tercih edilen yönsüz ve ağırlıklı graflarda; her düğüm bir cümleyi temsil etmekte olup Şekil 2 de graf olarak temsil edilmiş olan metin 7 cümleden oluşmaktadır. Ağırlıklı kenarlar cümleler arasındaki ortak kelime varlığını temsil etmektedir. Kenar ağırlıkları düğümler arası ortak kelime sayısı olarak alınmıştır. Örneğin, Şekil 2 de graf olarak temsil edilmiş metnin birinci ve ikinci cümlelerinin ortak kelime sayısı 2 dir.

3. GRAFLARIN YAPISAL ÖZELLİKLERİ

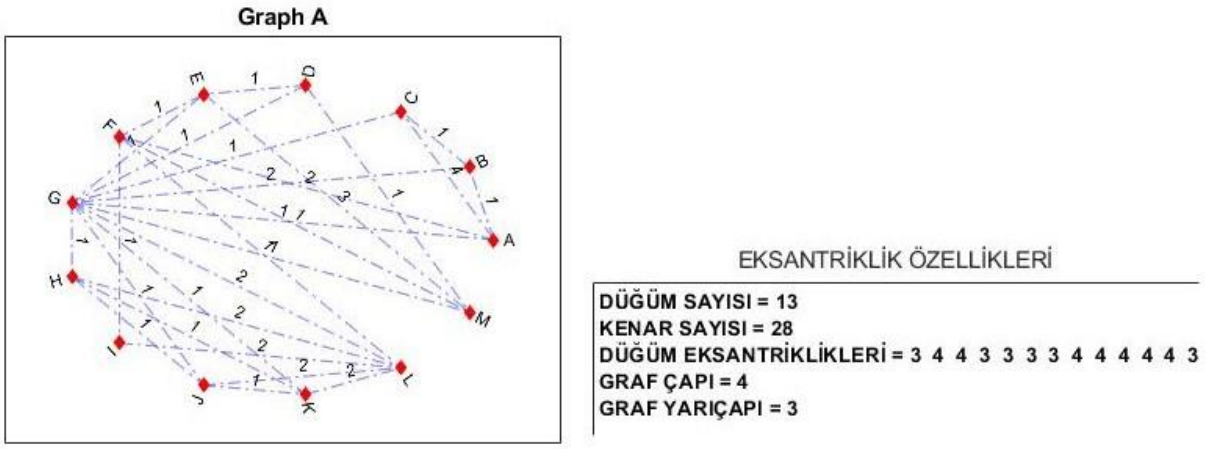
Düğümlerden ve kenarlardan oluşan grafların bazı yapısal özellikleri mevcuttur. Bir G grafında, bir düğümün **eksantrikliği** (eccentricity), bir düğümün diğer düğümlere olan maksimum en kısa yol mesafesidir (Peter D, 2004). Bir bağlı grafta, G grafındaki v düğümünün eksantrikliği, v ile G grafındaki diğer u düğümü arasındaki maksimum yol uzaklığıdır. Parçalı graflarda ise tüm düğümlerin eksantrikliği sonsuz olur. Graf düğümlerinin eksantriklik değerlerinin en küçüğü grafın **yarıçap**ını (radius), en büyüğü ise, grafın **çap**ını (diameter) vermektedir (Douglas BW, 2000).

Eccentricity: $\forall v \in V : \text{eccentricity}(v) = \max_{u \in V} |vu|$

Radius: $\text{radius}(G) = \min_{v \in V} \text{eccentricity}(v)$

Diameter: $\text{diameter}(G) = \max_{v \in V} \text{eccentricity}(v)$

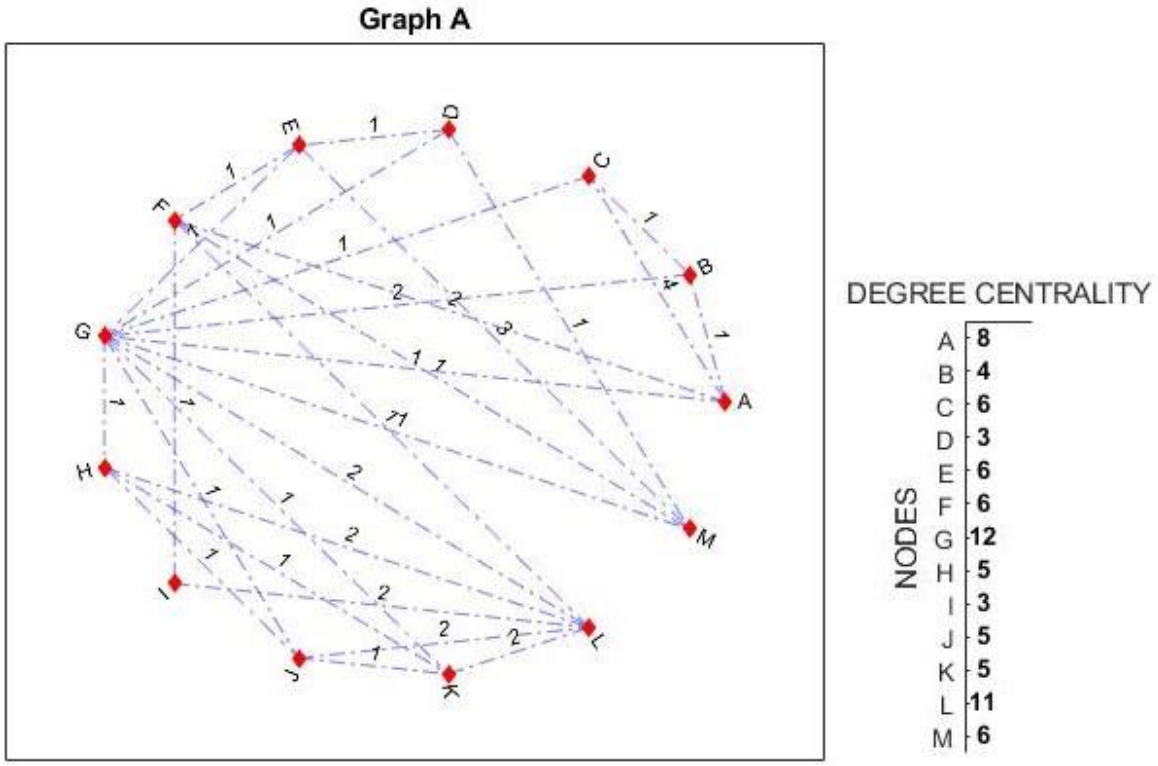
Şekil 3, 13 cümlelik bir metnin yani 13 tane düğümü olan A grafındaki düğümlerin eksantriklikleri düğüm sırasına göre verilip aynı zamanda, grafın yarı çapı ve çapı gösterilmiştir.



Şekil 3. Graf düğümlerinin eksantriklikleri

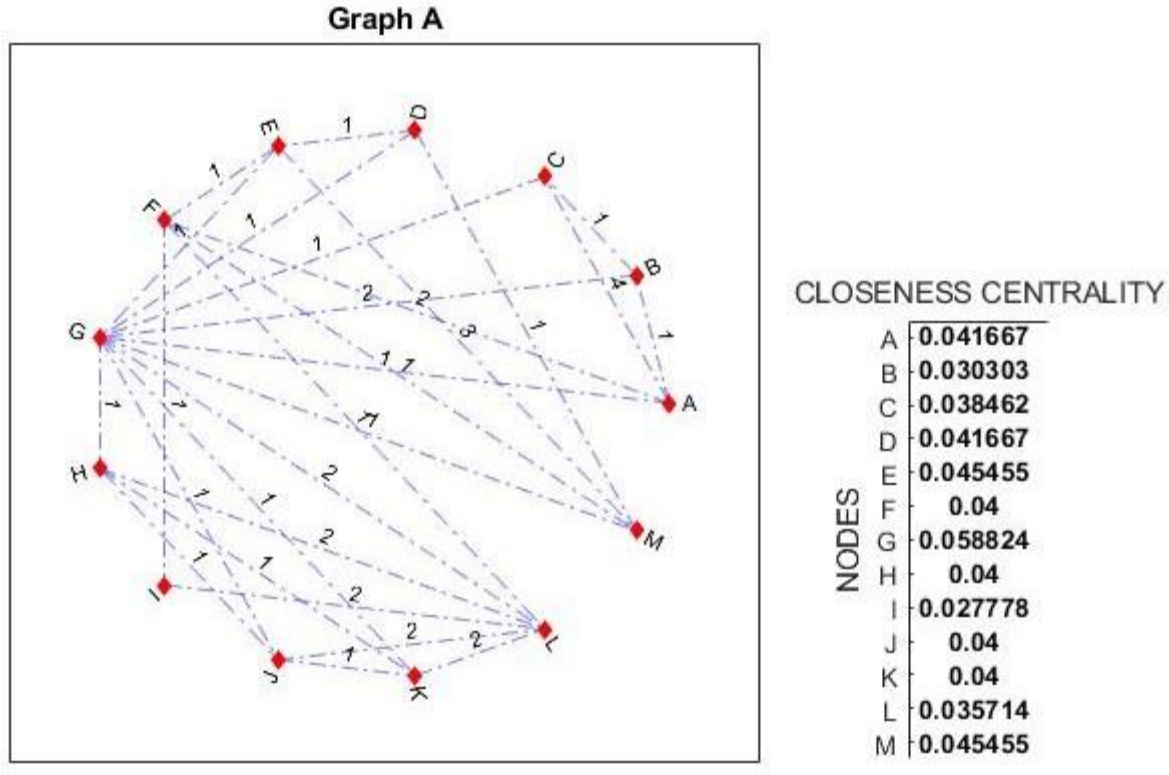
Graftaki bir düğümün graf **merkeziliği** (centrality), diğer herhangi bir düğüme olan maksimum mesafesinin tersidir (Amir A, 2015). Cümle merkeziliği kavramı ise Erkan G (2011) tarafından, bir cümlenin merkeziliği genellikle içerdiği kelimelerin merkeziliği açısından tanımlanır şeklinde ifade edilmiştir. Merkezilik, cümleler arasındaki benzerlik dikkate alınarak hesaplanır ve daha yüksek merkeziliğe sahip düğümler özet için daha önemli olarak kabul edilir. Bir düğümün graftaki diğer düğümlere bağlı olmasına dayanan çeşitli merkezilik türleri vardır ve hepsi graf için önemli olan düğümleri tanımlamak için kullanılır. **Derece merkeziliği** (degree centrality), bir düğümün derecesine, yani doğrudan ona bağlı olan kenarların sayısına bağlıdır. Yönlü bir grafta, bir düğümün derecesi, sosyal ağ analizinde sıklıkla yapıldığı gibi, gelen ve giden ayrıt (kenar) arasındaki bir ayrım temelinde hesaplanabilir (Zhang H, 2011).

n düğümlü G: = (V, E) grafı, v düğümünün derecesi (ona bağlı kenarların sayısı) deg(v) olsun bu grafta derece merkeziliği; C_D(v), v düğümü için: $C_D(v) = \frac{\deg(v)}{(n-1)}$, şeklinde hesaplanır (Zhang H, 2011). Şekil 4 te , A grafının düğüm merkezilikleri gösterilmiştir.



Şekil 4. A grafının düğüm merkeziliği

Yine graftaki bir düğümün **yakınlık merkeziliği** (closeness centrality), o düğümün diğer tüm düğümlere olan toplam mesafesinin tersidir (Amir A, 2015). Yakınlık merkeziliği, bir düğümün ağdaki diğer tüm düğümlere ne kadar yakın olduğunu gösterir. Düğümden ağdaki diğer düğümlere kadar olan en kısa yol uzunluğunun ortalaması olarak hesaplanır (Jennifer G, 2013). Yakınlık merkeziliği, her düğümün ağdaki konumunu, diğer ağ metriklerinden farklı bir bakış açısı ile ölçer ve her düğüm ile ağdaki diğer her düğüm arasındaki ortalama mesafeyi yakalar (Derek LH, 2019). Şekil 5 te , A grafının yakınlık merkezilikleri gösterilmiştir.

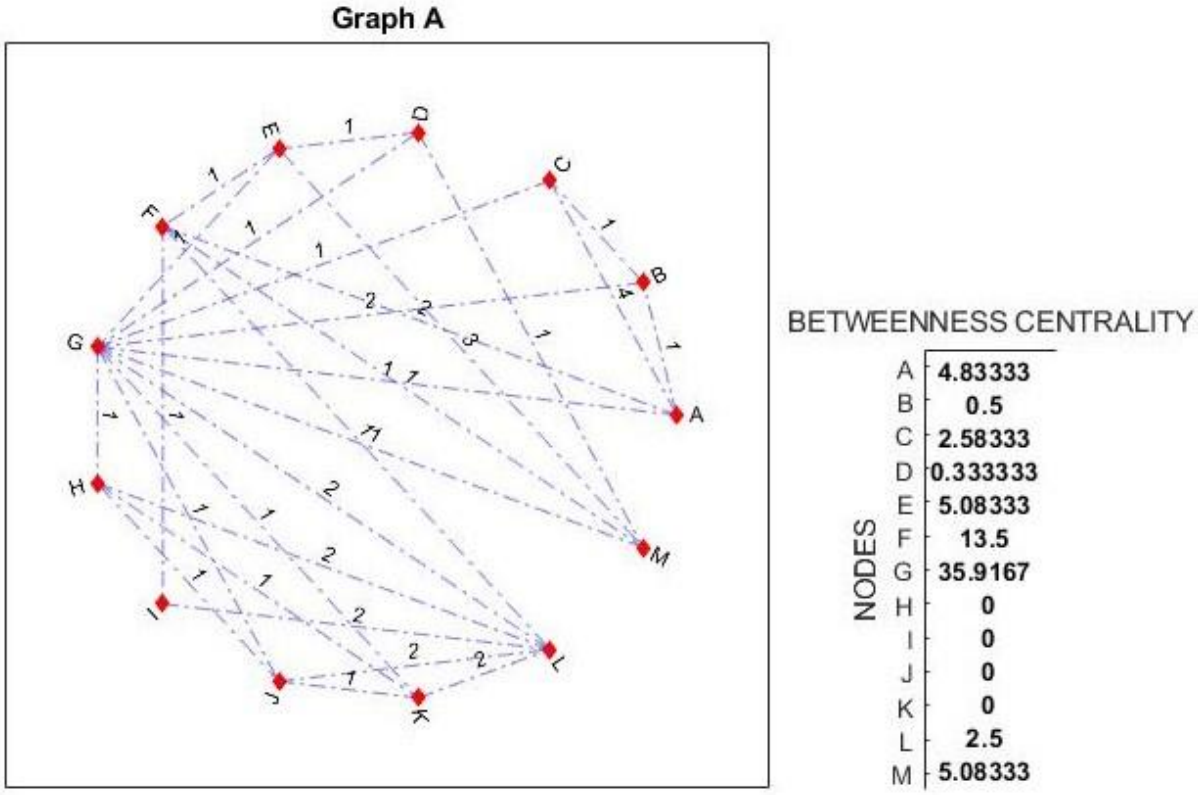


Şekil 5. A grafının yakınlık merkeziliği

Bir düğümün **arasındalık merkeziliği** (betweenness centrality), bu ara düğüme sahip olan en kısa yolların fraksiyonunu ölçer (Amir A, 2015). Arasındalık merkeziliği, bir düğüm bilgisinin öncelikle aralarındaki en kısa yollardan aktığı varsayımı altında, her bir düğüm noktası arasındaki bilgi akışı üzerindeki etkisinin bir ölçüsüdür. v düğümü için $C_B(V)$ arasındalık merkeziliği şu şekilde tanımlanır;

$$C_B(V) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

σ_{st}, uç düğümleri s ve t olan en kısa yolların sayısı, $\sigma_{st}(v)$ ise, v düğümünü içeren en kısa yolların sayısıdır (Sunil K, 2014). Şekil 6 da , A grafının arasındalık merkezilikleri gösterilmiştir.

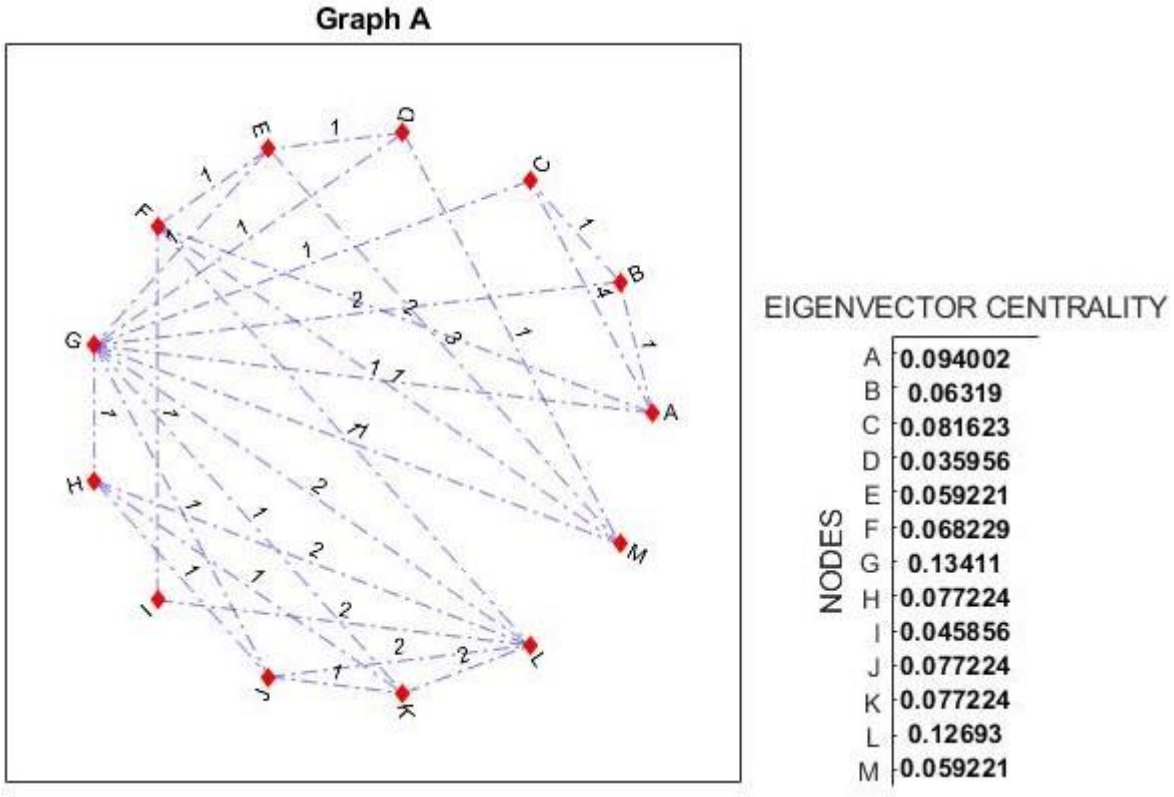


Şekil 6. A grafının arasındalık merkeziliği

Bir graftaki bir düğümün **özvektör merkeziliği** (eigenvector centrality), düğümün derecesinin yanı sıra komşularının derecelerinin bir ölçüsüdür. Özvektör merkeziliği (EVC), karmaşık ağlar alanında iyi bilinen bir merkeziyet ölçüsüdür. Bir ağ grafındaki köşelerin EVC' si, grafın komşuluk matrisinin ana özvektörüdür. Ana özvektör, grafın n-köşelerinin her biri için bir girişe sahiptir. Bir köşe için bu girişin değeri ne kadar büyük olursa, EVC' ye göre sıralaması o kadar yüksek olur (Meghanathan, N, 2015).

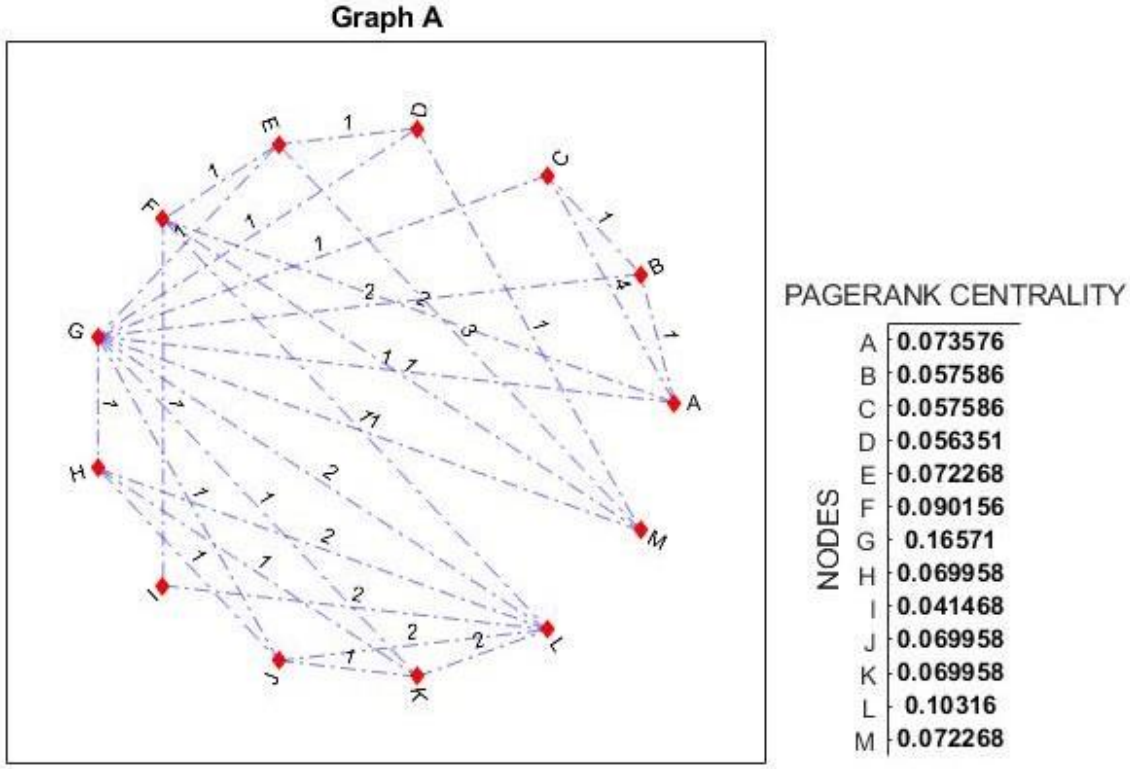
Özvektör merkeziliği, özellikle ağ içinde merkezi olan düğümlere bağlı düğümleri tercih eder. Böylece ağın tüm modelini dikkate alır. A, bir nxn benzerlik matrisini belirtsin. Daha sonra, i düğümünün özvektör merkezi x_i, A' nın en büyük öz değerine ait normalleştirilmiş özvektördeki i inci girişi olarak tanımlanır. Neden λ' nın en büyük özdeğer olduğunu ve x' in karşılık gelen özvektör olduğu incelendiğinde $Ax = \lambda x$ veya $x = \frac{1}{\lambda} Ax$, ve $x_i = \mu \sum_{j=1}^{n} a_{ij} x_j$ orantısallık faktörü ile $\mu = \frac{1}{\lambda}$ böylece x_i, kendisine bağlı tüm düğümlerin benzerlik puanlarının toplamıyla orantılıdır (Lohmann G, 2010).

Şekil 7 de , A grafının özvektör merkezilikleri gösterilmiştir.



Şekil 7. A grafının özvektör merkeziliği

PageRank merkeziliği, özvektör merkeziliğinin bir varyantıdır ve ağdaki bir düğümün rastgele bir yürüyüşle ziyaret edilme olasılığını yansıtır. PageRank, sözlük grafının her köşesine bir sıralama atar. Tüm köşeler aynı başlangıç sıralaması ile başlar ve daha sonra işaret ettikleri köşelere yinelemeli olarak dağıtılırken, kendilerine gösterilen köşelerden gelen sıralamaların toplamını alır (Michael WB, 2004). Şekil 8 de , A grafının pagerank merkezilikleri gösterilmiştir.



Şekil 8. A grafının pagerank merkeziliği

4. STANDART SAPMA

Standart sapma (standard deviation) değişkenliğin bir ölçüsüdür. Bir numunenin standart sapmasını hesaplandığında, numunenin alındığı popülasyonun değişkenliğinin bir tahmini olarak kullanılır. N skaler gözlemden oluşan rastgele değişken A vektörü için ; μ, A nın ortalaması olsun $\mu = \frac{1}{N}\sum_{i=1}^{N} A_i$ şeklinde ortalama hesaplandıktan sonra standart sapma, $S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|A_i - \mu|^2}$ ile hesaplanır.

Grafların elde edilen yapısal özelliklerinden benzerlik ölçüsü elde edebilmek için bu özelliklerin standart sapması alınmaktadır. Benzerlik ölçümü ise $\underset{\mu - S}{\longrightarrow}$ μ $\underset{\mu+S}{\longleftarrow}$ şeklinde alt ve üst sınırlar belirlenerek belli bir aralıkta hesaplanmış olur.

5. UYGULAMA

Uygulama geliştirme ortamı olarak mühendisler ve bilim adamları için özel olarak tasarlanmış güçlü bir programlama platformu olan MATLAB tercih edilmiştir. Geliştirilen uygulama ön kabuller ve 4 adımdan oluşmakta olup adımların sözde kodları tanımlanmıştır.

Ön kabuller:

Bir metnin cümle sayısıyla sınırlı büyüklükteki graflar söz konusudur. İki grafın kıyaslanabilmesi için yapısal olarak; bağlı graflar olması, kenar ve düğüm sayılarının aynı olması ön kabulleri mevcuttur.

Adım 1:

Metinler; cümleler düğümleri, kenarlar ise cümleler arası ortak kelime sayısını gösterecek şekilde yönsüz ve ağırlıklı graflarla modellenmiş ve komşuluk matrisleri çıkarılmıştır.

ALGORİTMA 1 : GRAF MODELLEME

```
1  Giriş – Metin A
2  For i ≔ 1 to A_CümleSayıs do
3        < Bul – cümleler arası ortak kelime sayısı >
4        < Oluştur - Komşuluk matrisi >
5  endfor
6  Graf(A_Komsuluk_Matrisi)
```

Adım 2:

Düğümlerden ve kenarlardan oluşan modellenmiş grafların; düğüm eksantriklikleri, graf çapı, graf yarıçapı, düğüm merkezilikleri, yakınlık merkezilikleri, arasındalık merkezilikleri, özvektör merkezilikleri ve pagerank merkezilikleri gibi yapısal özellikleri hesaplanmıştır.

ALGORİTMA 2 : GRAF YAPISAL ÖZELLİK ÇIKARMA

```
1  Giriş – Graf A
2  For i ≔ 1 to A_DugumSayısı do
3        Eksantriklik(A)
4        Merkezilik(A)
5  endfor
```

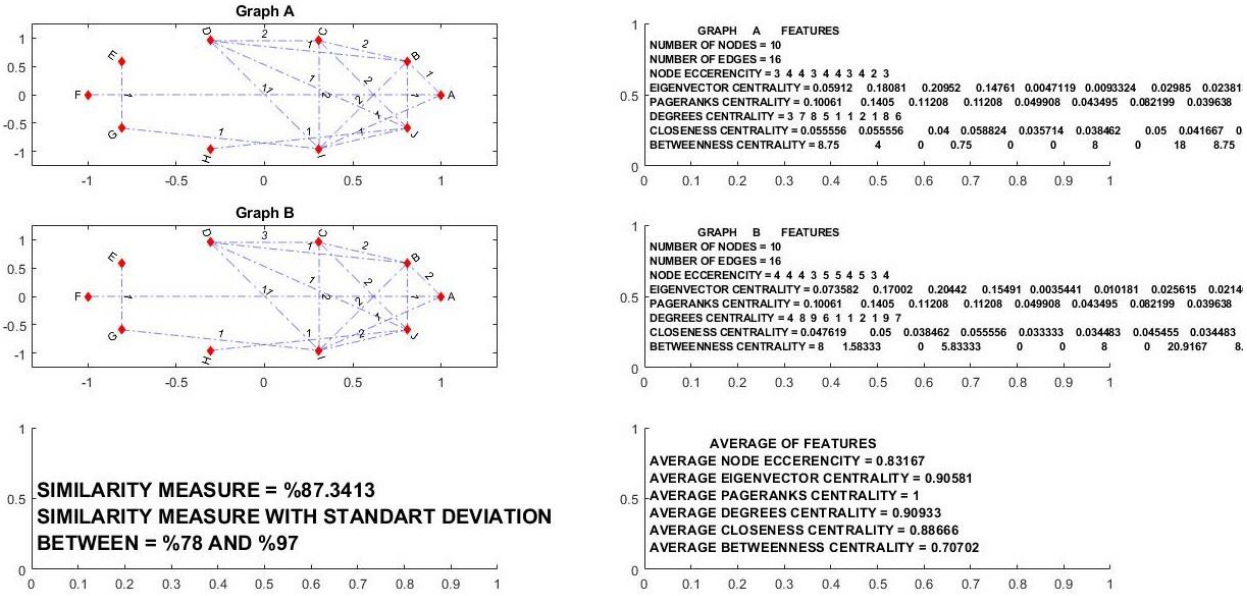
Adım 3:

Hesaplanan, graflara ait yapısal özelliklerin düğüm düğüm ortalamaları alınarak her düğüm için özellik değişimleri hesaplanmıştır.

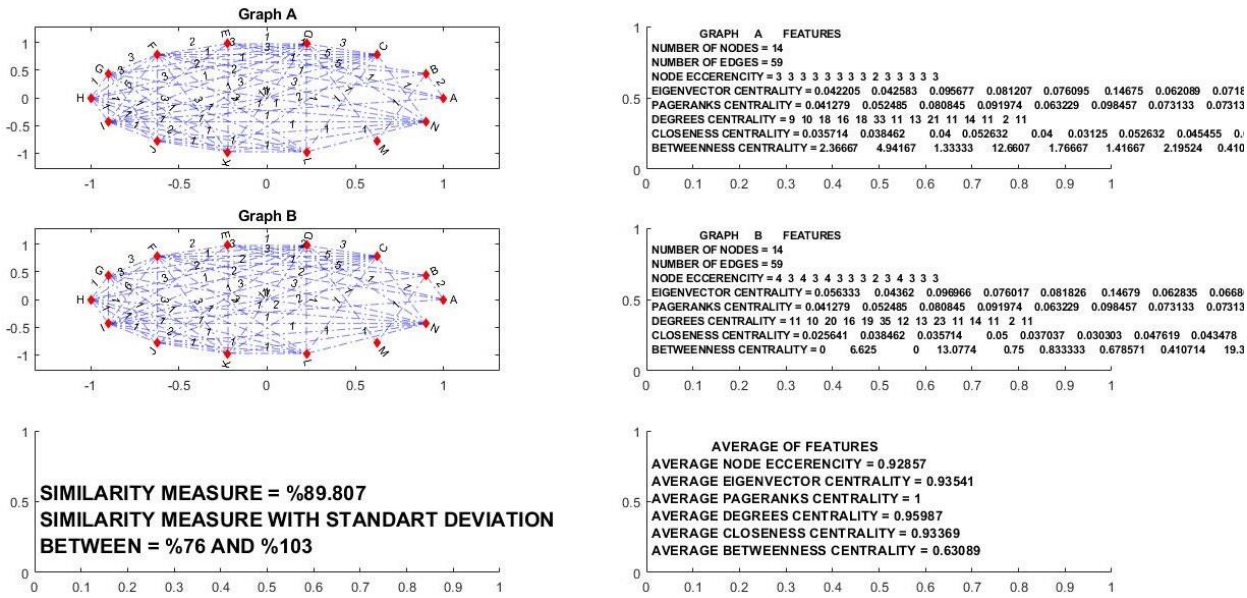
Adım 4:

Düğüm özellik değişimleri vektörel hale getirilip standart sapmaları elde edilerek benzerlik ölçüm aralığı hesaplanmıştır.

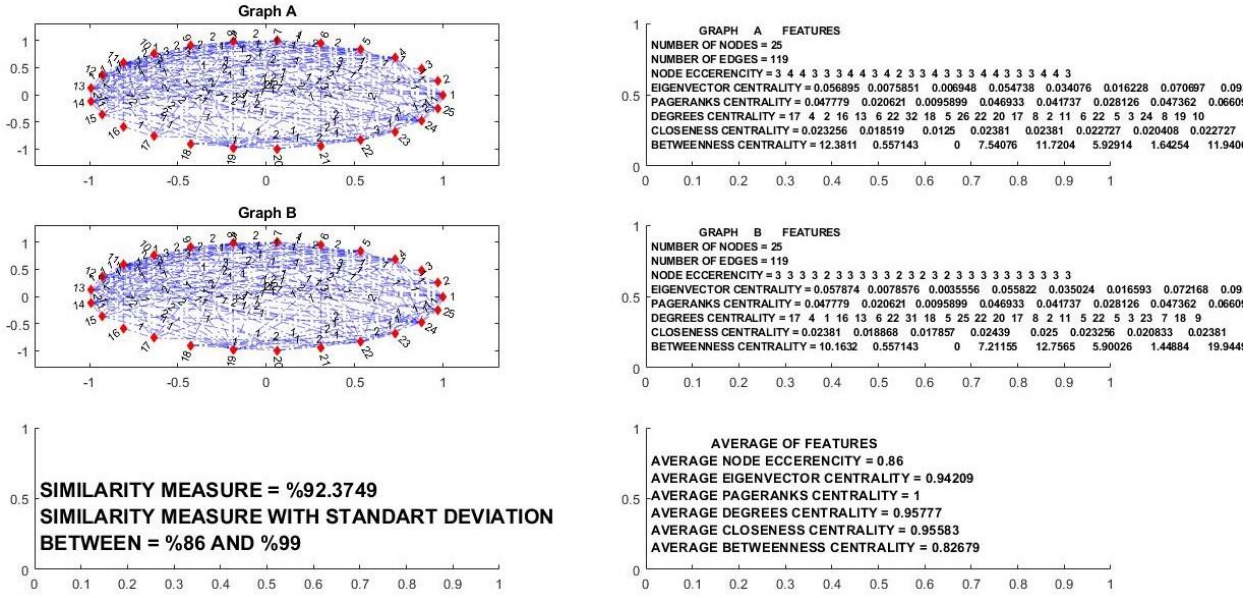
Şekil 9, 10, 11 ve 12 de örnek iki metnin graf benzerliği ile benzerlik ölçümü hesaplanıp Online Text Word Count uygulamasının sonuçları ile kıyaslanmıştır.



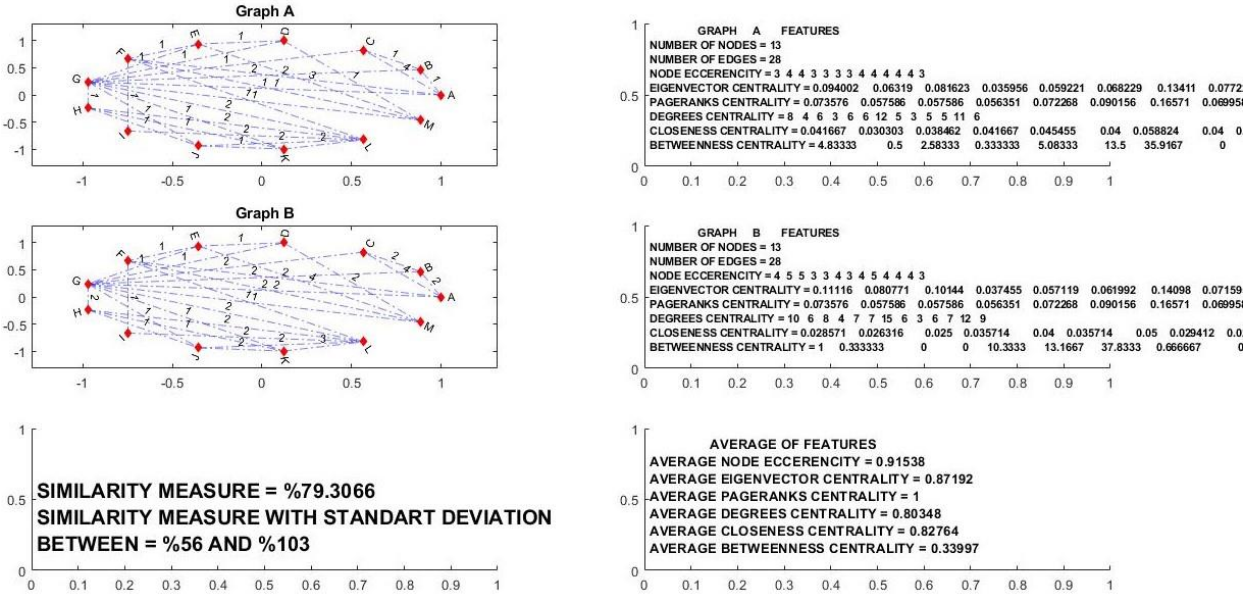
Şekil 9. İki metnin graf benzerliği ile benzerlik ölçüm uygulaması birinci örneği (Compare Text Online = % 84.36)



Şekil 10. İki metnin graf benzerliği ile benzerlik ölçüm uygulaması ikinci örneği (Compare Text Online = % 89.55)



Şekil 11. İki metnin graf benzerliği ile benzerlik ölçüm uygulaması üçüncü örneği (Compare Text Online = % 95.19)



Şekil 12. İki metnin graf benzerliği ile benzerlik ölçüm uygulaması birinci örneği (Compare Text Online = % 90.97)

SONUÇ

Geliştirilen yaklaşımın klasik benzerlik ölçümlerinden farkı, tüm metni değil de metni karakterize eden, cümleler arası ortak kelimelerin oluşturduğu grafların benzerliğinin ölçülmesidir. Doğrusal bir kıyaslamadan farklı olarak bir kelimenin diğer cümlelerde bulunup bulunmama durumuna göre bir benzerlik ölçüm mantığı geliştirilmiştir. Grafların benzerliği ölçülmeye çalışılırken de yukarıda özetlenen bilinen yöntemlerden farklı olarak, grafların yapısal özelliklerinin kıyaslanmasıyla düğüm bazlı hesaplanan yapısal özelliklerin graf kıyaslama da kullanılmasıdır.

İlerleyen çalışmalarda grafların, literetüre yeni girmiş yapısal özelliklerinin de hesaplanıp benzerlik ölçüm vektörüne dahil edilmesi ile başarımı arttırmak hedeflenmektedir.

KAYNAKLAR

Robin JW (1996) Introduction to Graph Theory Fourth edition, Addison Wesley Longman Limited, England.

John AB (1982) Graph Theory With Applications, Elsevier Science Publishing, USA.

Spizzirri, L. (2011). Justification and application of eigenvector centrality. Algebra in Geography: Eigenvectors of Network.

Peter D, Wayne G, Christine SS (2004). The Average Eccentricity of a Graph and its Subgraphs. Utilitas Mathematica. 65.

Douglas BW (2000) Introduction to Graph Theory (2nd Edition), Pearson, London

Amir A, Fabrizio G, Virginia VW (2015). Subcubic equivalences between graph centrality problems, APSP and diameter. In Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms (SODA '15). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1681-1697.

Erkan G, Dragomir R. (2011). LexRank: Graph-based Lexical Centrality As Salience in Text Summarization. Journal of Artificial Intelligence Research - JAIR. 22. 10.1613/jair.1523.

Zhang H, Fiszman M, Shin D, Miller CM, Rosemblat G, Rindflesch TC. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. Journal of biomedical informatics, 44(5), 830–838.

Jennifer G, (2013). Chapter 3 Network Structure and Measures, Analyzing the Social Web, Elsevier, Pages 25-44

Derek LH, Ben S, Marc AS, Itai H, (Available online 17 May 2019), Chapter 3 - Social network analysis: Measuring, mapping, and modeling collections of connections, Analyzing Social Media Networks with NodeXL (Second Edition), ,Elsevier, Pages 31-51

Sunil K, Balakrishnan K, Madambi J, Betweenness Centrality in Some Classes of Graphs, International Journal of Combinatorics, vol. 2014, Article ID 241723, 12 pages, 2014.

Meghanathan N, (2015). Use of eigenvector centrality to detect graph isomorphism, Computer Science & Information Technology (CS & IT), Academy & Industry Research Collaboration Center (AIRCC)

Lohmann G, Margulies DS, Horstmann A, Pleger B, Lepsien J, et al. (2010) Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. PLOS ONE 5(4): e10232.

Michael WB. (2004). Survey of Text Mining: Clustering, Classification, and Retrieval, Springer, USA

Vrotsou K, Johansson J, Cooper M, (2009). Activitree: Interactive visual exploration of sequences in event-based data using graph similarity. IEEE Transactions on Visualization and Computer Graphics, 15(6), 945-952.

Koutra, D., Parikh, A., Ramdas, A., & Xiang, J. (2011, December). Algorithms for graph similarity and subgraph matching. In Proc. Ecol. Inference Conf (Vol. 17).

Koutra, D., Vogelstein, J. T., & Faloutsos, C. (2013, May). Deltacon: A principled massive-graph similarity function. In Proceedings of the 2013 SIAM International Conference on Data Mining (pp. 162-170). Society for Industrial and Applied Mathematics.

Zhao, X., Xiao, C., Lin, X., & Wang, W. (2012, April). Efficient graph similarity joins with edit distance constraints. In 2012 IEEE 28th International Conference on Data Engineering (pp. 834-845). IEEE.

Fischer, A., Riesen, K., & Bunke, H. (2010, November). Graph similarity features for HMM-based handwriting recognition in historical documents. In 2010 12th International Conference on Frontiers in Handwriting Recognition (pp. 253-258). IEEE.

Naudé, K. A., Greyling, J. H., & Vogts, D. (2010). Marking student programs using graph similarity. Computers & Education, 54(2), 545-561.

Skvortsova, M. I., Baskin, I. I., Stankevich, I. V., Palyulin, V. A., & Zefirov, N. S. (1998). Molecular similarity. 1. Analytical description of the set of graph similarity measures. Journal of chemical information and computer sciences, 38(5), 785-790.

Runwal, N., Low, R. M., & Stamp, M. (2012). Opcode graph similarity and metamorphic detection. Journal in Computer Virology, 8(1-2), 37-52.

Sorlin, S., & Solnon, C. (2005, April). Reactive tabu search for measuring graph similarity. In International Workshop on Graph-Based Representations in Pattern Recognition (pp. 172-182). Springer, Berlin, Heidelberg.

Online Text Word Count, 2015-2019 COUNTWORDSFREE - Text Tools, https://countwordsfree.com

MATLAB, The MathWorks, Inc., 1994-2019, 1 Apple Hill Drive Natick, MA 01760-2098, www.mathworks.com