

PAPER DETAILS

TITLE: GERÇEK VE MODEL AGLARIN KARAKTERISTIK ÖZELLİKLERİNİN KARSILASTIRILMASI
- COMPARISON OF CHARACTERISTICS OF REAL AND MODELLED NETWORKS

AUTHORS: Sebnem AKAL

PAGES: 251-272

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/165849>

GERÇEK VE MODEL AĞLARIN KARAKTERİSTİK ÖZELLİKLERİNİN KARŞILAŞTIRILMASI

Şebnem AKAL*

Özet

Ağ analizi (network analysis) terimi uzun yıllara dayanan köklü bir geçmişe sahiptir. Ancak şu anda kullanıldığı biçimi bilgisayar teknolojilerindeki ilerlemelere paralel olarak ortaya çıkan, sosyal medya araçları sayesinde olmuştur.

Ağ analizinin, farklı alanlarda değişik uygulamaları bulunmaktadır. Günümüzde, sosyal ağların yanı sıra ekonomik ağlar, teknolojik ağlar, fiziksel ve biyolojik ağlar da mevcuttur ve bu ağların, benzer yöntemlerle analizleri yapılabilir. Bu durum, çok disiplinli bir ağ terminolojisinin doğmasına sebep olmuştur.

Bu çalışma, uzmanlık alanlarında ağ analizini uygulamak isteyen kişiler için, genel bir bilgi kaynağı olması amacıyla hazırlanmıştır. Çalışmada, ağların genel özellikleri farklı örnekler ile açıklanmış ve gösterilmiştir.

Çalışmanın birinci bölümünde, pek çok düğüm ve bağlantıdan oluşan gerçek ağlar tanımlanmıştır. Bu ağların incelenmesinde verilerin toplanması ve sınırlarının belirlenmesinin önemi vurgulanmıştır.

İkinci bölümde, gerçek ağların tanımı yapılarak gerçek ağların oluşturulmasında sınırların belirlenmesinden bahsedilmiştir. Üçüncü bölümde farklı ağ türlerini birbirinden ayıran önemli özellikler listelenerek açıklanmıştır. Dördüncü bölüm ise bilgisayarlarda çeşitli programlar ve algoritmalar yardımıyla üretilmiş olan sentetik ağlara ve ağ modellerine ayrılmıştır. Sentetik ağların, kullanım alanlarına ve özelliklerine bu bölümde yer verilmiştir.

Anahtar Kelimeler: Karmaşık Ağlar, Ağ Analizi, Ağ Modelleri

COMPARISON OF CHARACTERISTICS OF REAL AND MODELLED NETWORKS

Abstract

The term of “Network Analysis” has a long history based on many years. However its current usage format is determined by means of the social media tools which came up in accordance with the developments in computer technologies.

* Marmara Üniversitesi, İşletme Fakültesi, Almanca İşletme Enformatiği Bölümü, Öğretim Üyesi

Nowadays network analysis has various applications in different areas including social, economical, techonological, physical and biological networks which can be analyzed with similar methods. This situation has formed the basis of multi-disciplinary network terminology.

This study intends to be a general source of information for those who want to apply network analysis in their areas of expertise. In the study, general network features are explained and illustrated with different examples.

In the first part of the study, real networks consisting of many different nodes and connections are identified. The importance of gathering information and determining limits in network analysis are emphasized.

Second part includes real network definition and mentions determination of limits in formation of real network. The key features that separete different network types are listed and described in the third part. The fourth part includes synthetic networks and network models created by different computer programs and algorithms. Usage areas and features of the synthetic networks are also explained in this section.

Keywords: Complex Networks, Network Analysis, Network Models

1. Giriş

Ağ kavramı, 21. yüzyılın başından itibaren, pek çok bilim insanı tarafından “yeni bir bilim dalı olarak” tanımlanmaktadır. [1]–[3] Özellikle, enformasyon teknolojilerindeki gelişmelerin etkisiyle, 1990’ların sonundan itibaren ağlarla ilgili çalışmaların sayısında ciddi bir artış gözlenmiştir.

Çizge teorisi (*graph theory*) ve ağ analizi birbirleriyle önemli bağlantıları olan iki farklı alandır. Çizge, ağların matematiksel gösteriminde kullanılmaktadır. Her ikisi de matematiksel olarak aynı şekilde ifade edilir:

$$G = (V, E)$$

Ağlar ve çizgelerde (G) düğümler V (*vertices*), bağlantılar E (*Edges*) ile gösterilir. Ancak ağ analizi sadece çizgelerden ve matematik formüllerinden oluşmaz. Ağlar kendilerini oluşturan düğüm ve bağlantıların temsil ettikleri değerlere göre farklı konularla ilgili olabilir: Düğümleri insanlar, bağlantıları ilişkiler olan sosyal ağlar; düğümleri bilgisayarlar, bağlantıları kablolu ya da kablosuz bağlantılar olan bilgisayar ağları ilk akla gelen örneklerdir. Bu örnekler çoğaltılabilir. Fizikten biyolojiye, matematikten sosyal bilimlere, politikaya kadar pek çok farklı konuda ağ analizinden yararlanılmaktadır.

Ağ analizinin çeşitli alanlarda uygulamaları bulunmaktadır. Analizin konusu gerçek ağlar olsa da, gerçek ağlar kullanılmadan, belirli parametrelerle yaratılan sentetik ağlarla da incelemeler yapılabilir. Güvenilir sonuçlar elde etmek için ağın yapısının tanınması, özelliklerinin iyi bilinmesi gerekir. Bunun için tesadüfilik, entropi, merkezilik ölçüleri, düğüm derece dağılımı, çap ve yarıçap, kümeleme ve korelasyon katsayısı gibi bazı metrik değerler hesaplanmalıdır. Bu parametrelere dayanarak gerçek ağları temsil eden sentetik ağlar yaratılabilir ve enformasyonun, sinyallerin ya da fikirlerin ağ üzerinde yayılması incelenebilir.

Düzenli, tesadüfi, ölçek-bağımsız ağlar ve küçük-dünya ağları parametrelerdeki çeşitli değişiklikler sonucunda ortaya çıkan ağ modelleridir. Bu makalede, parametrelerdeki değişikliklere göre hangi ağ türlerinin oluştuğu ve bu türlerin hangi çalışmalarda kullanıldığı tartışılacaktır. Aynı zamanda gerçek ağların özelliklerinden yola çıkılarak benzer özellikleri taşıyan sentetik ağların üretilmesinde de bu dört farklı ağ modelinden yararlanılmaktadır.

2. Gerçek Ağların Tanımlanması ve Sınırlarının Belirlenmesi

Düğüm ve bağlantıların sayısı çok fazla olmadığı sürece, ağ analizi karmaşık bir yapı değildir. Ancak, enformasyon teknolojilerindeki gelişmelere paralel olarak düğüm ve bağlantıların sayısı çok büyük boyutlara ulaşmıştır. Düğüm ve bağlantıların sayısı fazla değilse, çeşitli görselleştirme yazılımları kullanılabilir, ancak sayı arttıkça, görselleştirme araçlarını kullanmak imkansız hale gelir. Bunun yerine hesaplamalarda, matrisler ve algoritmalarından faydalanmak yerinde olacaktır.

Örneğin, bir kişiye ait kişisel iletişim ağı (*personal contact network*) önceleri 100 ya da en fazla 200 düğümden oluşurken, günümüzde *Twitter*, *Facebook* benzeri sosyal medya kullanımının yaygınlaşması ile bu sayı artmış, bilgisayar yardımı olmadan incelenemez boyutlara ulaşmıştır. Eylül 2012 itibarıyla *Twitter*'in 500 milyondan fazla, *Facebook*'un 1 milyar aktif kullanıcısı bulunmaktadır. [4]

Bu büyüklükteki ağlara, düğümlere ait öznitelik değerleri de eklendiğinde, ağ çok büyük boyutlara ulaştığı için çizge ile ifade edilemez. Sadece bazı parametreler yardımıyla ağ üzerinde değerlendirme yapılabilir.

Karmaşıklık (*complexity*) dolayısı ile, her şeyin birbiri ile bağlantılı olduğu gerçek hayatta, bu kadar geniş ağların oluşturulması ve incelenmesi mümkün değildir.

Güvenilir sonuçlara ulaşmak için ne kadar büyüklükte bir ağ gereklidir? Bu karar nasıl ve kim tarafından verilecektir? Bu sorulara cevap bulabilmek için ağ kavramını tanımak ve çeşitli özelliklerini ölçebilmek gereklidir. Örneğin, dünyanın kendisi aslında sınırları belirlenmemiş büyük bir ağıdır: Düğüm sayısı milyonlarca ya da milyarlarca olabilir. 2012 yılı Haziran ayı verilerine göre dünya nüfusu 7 milyardan fazladır. [5] Düğüm sayısı 7 milyar olan bir sosyal ağ kaç adet bağlantıya sahiptir? Sorunun cevabını yaklaşık olarak hesaplamak için her düğümün bağlantı sayısının *Dunbar sayısına* eşit olduğunu varsayılmaktadır. Bir kişinin ortalama sosyal çevresi (yani ortalama bağlantı sayısı k) *Dunbar*'ın çalışmalarına göre 150'dir. [6] Düğümlere ait hiç bir ek öznitelige yer verilmese de, 7 milyar düğüm ve her düğüme ait 150 bağlantı ile ortaya çıkan ağ ise fazla büyük bir ağıdır. Oysa her kişinin 150 bağlantısı yoktur. Gerçek dünyada çok az sayıda kişinin çok sayıda bağlantısı varken, çok sayıda kişi az sayıda bağlantıya sahiptir. Bu ağlar seyrek ağlar olarak tanımlanır. Yayılma modellerinde de bu durum ayrıntılı olarak incelenmiştir. [7]–[10] 1982 Hastalıkların, fikirlerin ya da enformasyonun yayılması sadece yayılan hastalık, fikir ya da enformasyonun özelliklerine bağlı değildir, ağın yapısal özellikleri de yayılma üzerinde çok etkilidir. [11, pp. 645–647]

Ayrıca ağ, merkezi ya da yönetici bir düğüm içermemektedir. Düğüm sayısı (n) olan bir ağda kişilerin ortalama bağlantı sayısı (k) ile ifade edilsin.

Maksimum bağlantıya sahip olan düğümler dahi, hiç bir zaman düğümlerin tamamı ile bağlantılı olamaz. Yani, örnekteki gibi dünya üzerinde hiç bir düğümün 7 milyar sayıda bağlantısı yoktur. Zaten, seyrek ağ özelliği taşıyan ağların merkezi bir yöneticisinin olması tanım itibarıyla da mümkün değildir.

Tüm düğümlerle bağlantılı tek bir merkezi düğüm olmadığına göre, hastalıklar ya da enformasyon bu kadar hızlı nasıl yayılmaktadır? Bu sorunun cevabı ağ içerisinde pek çok farklı grubun olmasıdır. Bu gruplar ise, yukarıda sözü edilen bağlantı sayısı fazla olan düğümler tarafından birbirine bağlanır. Bir ağda, bu durumun varlığı ve derecesi kümeleme katsayısı ile ölçülmektedir.

Gerçek ağlar, topolojik ve yapısal olarak birbirinden farklı özelliklere sahip olabilir. Ağları yapısal özelliklerine göre ayırmak ya da sınıflandırmak için bazı hesaplamalar yapılabilir. Bu hesaplamalar merkezi düğümlerin belirlenmesi, derece sayısı gibi sadece düğümlerle ilgili ölçümler olabileceği gibi; derece dağılımlarının belirlenmesi, kümeleme katsayısı gibi ağın tümüne ilişkin ölçümler de olabilir. Bu ölçümlerden bir kısmı, ağ modellerini belirlemekte de kullanılmaktadır. Tüm bu özellikler ilerleyen bölümlerde ayrıntılı olarak anlatılmıştır. Ayrıca, bu tür ağ istatistikleri, ağların genel özelliklerini tanımlamaya, farklı ağ türlerini birbirinden ayırmaya ve kullanılabilir algoritma ve uygulamalar oluşturmaya hizmet edebilir.

3. Gerçek Ağların Yapısal Özellikleri

Bir ağın yapısal özellikleri ağın topolojisi ile ilişkilidir. Ağın topolojik özelliklerinden etkilenen bazı ölçümler ağın yapısal özelliklerini belirlemede de yardımcı olur.

Ağları, üretilmiş ağlar ve gerçek ağlar olmak üzere iki gruba ayrılabilir. Üretilmiş ağlar, yani sentetik ağlar ya da diğer bir ifade ile idealize edilmiş ağlar, belli ölçüler göz önünde tutularak yapay bir şekilde oluşturulmuş ağ yapılarıdır.[12] Sentetik ağlar sayesinde, ağ analizinde parametre değişiklikleri gözlenerek daha somut araştırmalar yapılabilmektedir.[13] Daha sonra, elde edilen bulgular gerçek ağlarda kullanılmaktadır.

Yayınlanan çeşitli çalışmalarda, gerçek ağların bazı özelliklerinden bahsedilmektedir.[2], [14]–[21] Buna göre gerçek dünyadaki ağlar genellikle şu özellikleri taşımaktadır:

1. Ağ, **seyrek ağ** (*sparse network*) özelliği gösterir.
2. Düğüm derece dağılımı, genellikle **güç yasası dağılımına** (*power-law distribution*) uymaktadır.
3. **Ortalama yol uzunluğu** kısadır. Bu durum, **küçük dünya teorisi** (*small world theory*) diye adlandırılan fenomenin bir sonucudur.
4. Düğümler arasındaki **kümeleme katsayısı** (*clustering coefficient*) yüksektir.
5. Son olarak, benzer özelliklere sahip düğümler arasında pozitif **seçici bağlantılar** (*selective linking*) vardır.

Tüm ağ örnekleri söz konusu olduğu için yapılan genellemenin sınırları ve istisnalarının olduğunu vurgulamak yerinde olacaktır. Önceki bölümde ifade edilmiş olduğu gibi, ağ kavramı bir soyutlamadır ve pek çok disiplinde uygulama örneklerini görmek mümkündür. Maalesef bu nedenle

yukarıdaki liste, gerçek ağların tamamında aynı şekilde geçerli değildir. Örneğin, sosyal ağlarda derece dağılımı güç yasasına uyum gösterirken, biyolojik ağlar da bu özellik görülmez. İkinci olarak, gerçek ağlarda, genellikle pozitif seçici bağlantılara rastlanmaktadır. Ancak bazı protein ilişkilerinin incelendiği metabolik ağlarda ya da av-avcı ilişkilerinin gösterildiği ağlarda negatif seçici bağlantılar gözlenmiştir. Bu durum ilgili bölümlerde örnekler verilerek ayrıca açıklanacaktır.

Bu özelliklerin anlaşılması için içerdikleri terimlerin de anlaşılması gereklidir. Her ne kadar ağların oluşturulma mantığı çok basit olsa da (düğüm ve bağlantılardan oluşur) ağ terminolojisi karmaşık bir yapıya sahip olup anlaşılması için pek çok matematiksel terimin iyi bilinmesi gerekir. Aşağıdaki bölümde, bu kavramlardan en önemlileri ayrıntılı açıklanacaktır. Daha detaylı bilgi için 2000’li yılların başlarından itibaren yayımlanmış olan “ağ bilimi” ile ilgili önemli çalışmalar incelenebilir. [1]–[3]

3.1. Düğüm Derece Dağılımı

Düğüm derece dağılımı, ağ istatistikleri içerisinde en eskilerinden ve önemlilerindendir. Bir düğümün derecesi her bir düğüme ait toplam bağlantı sayısıdır. Düğüm derece dağılımı ise her bir düğümün derecesinin ağdaki toplam bağlantı sayısına bölünmesi ile hesaplanmaktadır.

Şekil 1’deki tabloda, düğümlerin dereceleri ve frekans dağılımları gösterilmiştir. Toplam 10 düğümden oluşan bu küçük örnekte birinci tabloda, her bir düğüme ait derece (k_i) değerleri gösterilmiştir. Yandaki sütundaki ikinci tabloda ise dereceler için frekans dağılımları ve olasılık dağılımları bulunmaktadır. $P(k_i)$ ise tesadüfi olarak seçilen bir düğümün derecesinin k_i ’ye eşit olma olasılığını göstermektedir.

En çok bağlantı sayısı 5, en az bağlantı sayısı 3’tür. Toplam 3 adet bağlantısı olan 3, 4 bağlantısı olan 4 ve 5 bağlantısı olan 3 düğüm bulunmaktadır. Olasılık olarak bakıldığında bu çizgedeki her hangi bir düğümün 4 adet bağlantısının olma olasılığı %40 olmuştur.

derece (k_i)	frekans dağılımı
1	3
2	3
3	4
4	4
5	5
6	5
7	4
8	5
9	4
10	3

derece (k_i)	frekans dağılımı	$p(k_i)$
1	0	0.00
2	0	0.00
3	3	0.30
4	4	0.40
5	3	0.30
Toplam	10	0.00

Şekil 1: Düğüm derece dağılımının hesaplanması

Derece dağılımı, *histogram* yardımı ile de gösterilebilir. Eğer, düğüm sayısı çok fazla ise logaritmik ölçekten yararlanmak daha yerinde olacaktır. Daha fazla örnek için, 2003 yılında Newman

tarafından hazırlanmış olan, ağ özelliklerinin karşılaştırmalarıyla birlikte ayrıntılı bir şekilde anlattığı çalışma incelenebilir. [20, p. 14]

Gerçek ağların derece dağılımları genellikle güç yasasına uymaktadır. Dördüncü bölümde ayrıntılı anlatılacak olan ağ modelleri arasında bulunan ölçek-bağımsız ağlarda benzer derece dağılımına sahiptir. Tesadüfi ağların derece dağılımı ise *poisson* ya da *binom* dağılıma uymaktadır.

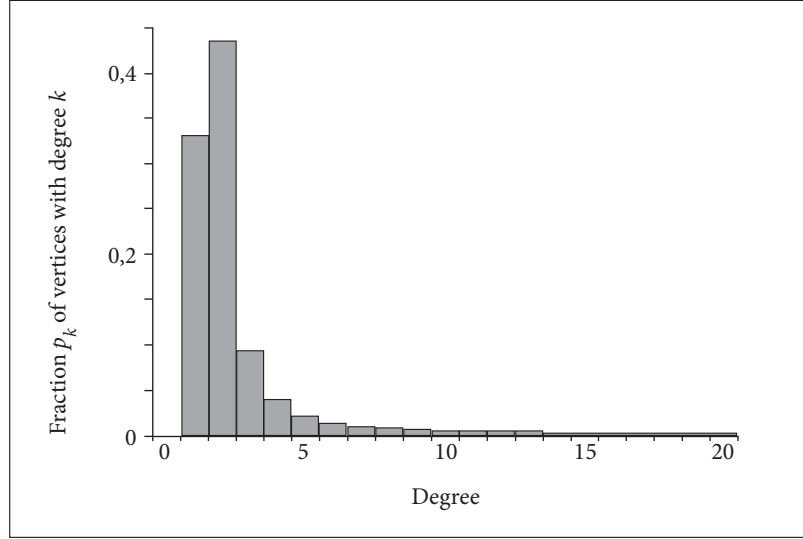
Pareto'nun, kanonik güç dağılımı olarak tanımladığı yasanın temel prensibi Zipf'in sloganı çok iyi açıklamakta ve bu gün yaygın bir şekilde kullanılmaktadır: "Zengin daha zengin olur!" Az sayıda kişi servetin büyük kısmını, çok sayıda kişi ise geri kalanını paylaşmaktadır.[21, p. 176] Zipf'in bir diğer vurgusu da sistemin sürekli genişlemesidir. Bu oransal genişleme ilerleyen yıllarda ağ analizinde "ölçek bağımsız" isimli ağ modelinin tanımlanmasında da kullanılmıştır.[22]

Güç yasasının, ağ analizinde kullanılması 1999 yılında *Nature* dergisinde yayınlanan iki farklı çalışmaya dayanmaktadır.[14],[15] Bu çalışmalardan biri Albert, Jeong ve Barabási tarafından WWW'deki (*World Wide Web*) düğümlerin derece dağılımlarının incelendiği çalışmadır. Derece dağılımlarının, güç yasasına uygun olduğu ilgili çalışmalarda vurgulanmaktadır.

Düğüm derece dağılımının güç yasasına uygun olmasının anlamı nedir? Bu sorunun cevabı bir ağda enformasyonun, hastalıkların yeni fikirlerin nasıl yayıldığına anlaşılmamasını da sağlayacaktır. Barabási'nin çalışmasında WWW incelenmiştir. WWW'in toplam düğüm sayısını tahmin etmek çok güçtür. Her gün yeni dokümanlar ve bağlantılar eklenmekte ve sürekli genişlemektedir. Bu nedenle tamamının haritalandırılması mümkün değildir. Ancak merkezi bir düzenleme olmamasına, kişiler web sayfalarını yaratırken seçimlerinde tam olarak özgür olmalarına rağmen bazı düğümler diğerlerine göre daha çok tercih edilmektedir. Bu durum Barabási ve çalışma arkadaşları tarafından Web'in sürü mantığının ve kendini düzenleyen yapısının bir göstergesi olarak açıklanmıştır.[23] Bu karmaşık ağ, kendini düzenleyen yapıya sahip bir sistemdir. Bunun sonucunda bazı düğümler daha çok tercih edilmekte ve bu düğümler sayesinde ağ birbiri ile olağanüstü düzeyde bağlantılı bir yapıya dönüşmektedir. Tahmin edilene göre 8×10^8 düğüme sahip olan bu devasa ağda tesadüfi olarak seçilen düğümler arasındaki ortalama uzaklık sadece 19 bağlantıdır. Böyle büyük bir ağda bir düğümden diğerine 19 adımda ulaşabilmek demek, bir bilginin de bir diğerine ne kadar hızlı iletebileceğinin de göstergesidir.

Bu durum, ağın sürekli genişlemesinin ve *tercihli büyümenin* bir sonucudur. Ağa sürekli bir şekilde yeni düğümler eklenerek ölçek bağımsız bir düzende büyüyen bir ağ yapısı ortaya çıkmaktadır. [25] Tercihli bağlantı özelliği nedeni ile de ağın sürekli bağlantılı olarak kaldığı ifade edilmiştir. Yeni bir buluş olmasa da, ağların yapısının güç yasası ile açıklanması ağ analizi çalışmalarında çığır açan bir yenilik olmuştur.

Internet ve aktör-film veri tabanı incelendiğinde de düğüm derece dağılımlarının güç yasasına uyduğu görülmüştür.[26, p. 306] Şekil 2'deki *histogram* Internet'in derece dağılımını göstermektedir. Pareto diyagramı olarak da bilinen histogram derece dağılımını göstermekte sık kullanılan bir ifade şeklidir. Yaklaşık 20.000 düğümün bulunduğu örnekteki veri setinde az sayıda düğüm ağın 12'si ile bağlantılıdır. [27, p. 188]



Şekil 2: Internet histogram [27, p. 188]

Gerçek dünyada gözlenen ağların pek çoğu güç yasasına uygun olsa da, tümü bu dağılımın özelliklerine uygun değildir. Yoğun bağlantılı (*densely connected*) ya da tam bağlantılı (*completely connected*) ağlar güç yasası dağılımına uymazlar. Eğer bir ağda, her bir düğümden (yaklaşık olarak) diğer tüm düğümlere bağlantı varsa bu durumda ağ yoğun ağ (*dense network*) olarak adlandırılmaktadır. Pek çok araştırmacı tarafından gerçek ağların dağılımının güç yasasına uygunluk gösterdiği ifade edilse de bu durum yoğun ve tam bağlantılı gerçek ağlar için geçerli değildir. [28]

Bundan dolayı, güç yasasına tek başına gerçek dünyadaki ağların tümünü anlamak ve çözümlmek için yeterli değildir.[27, p. 188] Bunun yanı sıra, diğer istatistik bilgilerin de incelenerek yorumlanması yerinde olacaktır.

3.2. Ortalama Yol Uzunluğu

Bir ağın sahip olduğu ortalama yol uzunluğu, ağ içinde bir düğümden diğer tüm düğümlere ulaşmak için atlamak gereken ortalama en kısa adım sayısını göstermektedir. Ortalama yol uzunluğu değişik ağ modelleri için farklılık göstermektedir.

Toplam düğüm sayısının (n) olduğu bir ağda (i) ve (j) düğüm ve bağlantılar olmak üzere, $d(i,j)$ düğüm ve bağlantılar arasındaki toplam en kısa yol uzunluğunu (*geodesic*) göstermektedir. Ortalama yol uzunluğu (l) aşağıdaki şekilde hesaplanabilir:

$$l = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d(i,j)$$

Ortaya çıkış noktası ise, Milgram'ın küçük dünya deneyi (*small-world experiment*) isimli çalışmasına dayanmaktadır. [29, pp. 61–63] Carlos Pinheiro, 2011 yılında yayımlanan kitabında bu konuya değinerek; yıldız, çizgi ve daire şeklindeki çizgelerde düğümler arasındaki mesafeyi

(adım sayısını) karşılaştırmıştır. [30, p. 70] Ayrıca bu değerin yakınlık merkeziliği (*closeness centrality*) olarak da bilinen bir merkezilik ölçüsü ile de hesaplanması mümkündür.

Gerçekten de ortalama yol uzunluğunun kısa olması, küçük-dünya fenomeninin bir göstergesidir. Küçük dünya fenomeninin gözlemlendiği ağlarda, çizge büyüdükçe ortalama yol uzunluğu kısalmaktadır. [31]

Ortalama yol uzunluğu, bir ağdaki enformasyonun yayılma gücü ve yönü hakkında bilgi sahibi olmak açısından önemli bir göstergedir. Eğer ortalama yol uzunluğu çok kısa ise bu durum, bazı kişi ya da gruplara enformasyonun ulaşmadığının bir göstergesi de olabilir. [32, p. 17] Ancak, diğer göstergelere bakılmadan tek başına bu konuda bir yorum yapmak doğru olmayabilir. Kümeleme katsayısına da bakılması yorumlamayı daha anlamlı hale getirecektir.

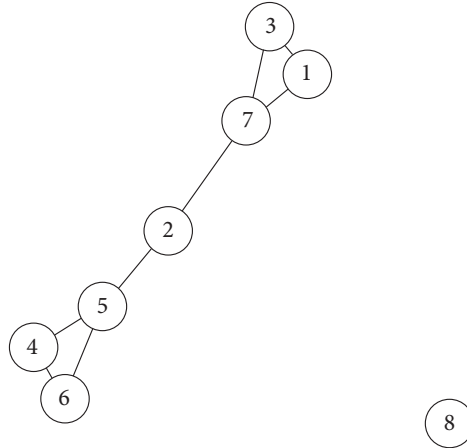
3.3. Kümeleme Katsayısı Büyüklüğü

Pek çok ağda gözlenen bir diğer özellik de büyük kümeleme katsayısına sahip olmalarıdır. Kümeleme katsayısı bir ağın düğümleri arasındaki gruplaşmanın derecesini ölçmek için kullanılan bir ölçüdür.

$$C_i = \frac{\text{Üçgen şeklinde (triangle) bağlı düğüm sayısı}}{\text{Birbirine bağlı üçlü düğüm sayısı (triples)}}$$

$$C = \frac{1}{n} \sum_i C_i$$

Bu kavramları bir örnekle açıklamak gerekecektir. Şekil 3’de görüldüğü gibi 4, 5 ve 6 numaralı düğümler ile 1, 3 ve 7 numaralı düğümler birbirlerine üç farklı bağlantı ile bağlıdır. Pay kısmında kastedilen bu tür bağlantılardır. Payda da görülen “üçlü bağlantı” ise şekilde 2, 5 ve 7 numaralı bağlantılar arasında görülmektedir. Her bir düğüm için bu formül uygulanır ve bu toplam değerin (n) sayıda düğüme oranı hesaplanırsa o çizgeye ait kümeleme katsayısı hesaplanabilir.



Şekil 3: Kümeleme katsayısının hesaplanmasında kullanılan üçlü ve üçgen şeklindeki bağlantılar

Şekil 3’de görülen 8 numaralı düğüm ise izole bir düğümdür ve hiç bir bağlantısı yoktur. Eğer bir ağda, bu tür izole düğümlerin sayısı fazla ise, o ağın kümeleme katsayısı düşük olacaktır.

Kümeleme katsayısı 0 ile 1 arasında değişir. Gerçek ağlarda, zaman içinde ağın büyüklüğü değişse bile kümeleme oranı sabittir, değişim göstermez.

Kümeleme katsayısı bir ağdaki enformasyonun yayılma düzeyi hakkında da fikir verebilir. Başka bir ifadeyle, bir grup içindeki düğümlerin birbirini ne kadar iyi tanıdığı kümeleme katsayısı ile hesaplanabilir. [33, p. 17] Gruplar ne kadar kalabalıksa ya da düğümler arasında ne kadar çok bağlantı varsa enformasyon o derece hızlı yayılır. Yerel enformasyonun yani grup içinde enformasyonun yayılma düzeyi kümeleme katsayısının büyüklüğüne bağlıdır.

Ortalama yol uzunluğu ile birlikte değerlendirildiğinde, kümeleme katsayısı bir ağın küçük-dünya fenomeninin özelliklerini taşıyıp taşımadığının anlaşılmasına yardımcı olur.

3.4. Seçici Bağlantı

Ağın en önemli özelliklerinden birisi de bağlantıların sürekli yenilenmesidir. Bazı düğümler arasındaki bağlantılar kaybolurken, farklı düğümler arasında yeni bağlantılar ortaya çıkmaktadır. Bu devinim, tesadüfi değildir. Bağlantı gelişigüzel iki düğüm arasında oluşmaz. Düğümler, bağlantılarını belirlerken, genellikle benzer özelliklerine göre seçim yaparlar. Bu durum, seçici bağlantı (*selective linking*) olarak tanımlanmaktadır.[20] Farklı uzmanlık alanlarında aynı durumu ifade etmek için kullanılan kavramlar da değişiklik gösterir. En sık karşılaşılan ve aralarında bazı nüanslar bulunsun da, aynı anlama gelen üç kavram olan seçicilik (*assortativity*), homofili (*homophily*) ve modülerlik (*modularity*) zihinlerde anlam karmaşası yaratmaktadır. [34]

Üç kavramı birbirinden ayıran küçük farklar vardır. [35] *Homofili*, sosyal ağ analizi çalışmalarında kullanılan bir kavramdır. [36]–[39] Sosyal ağlarda, bireyler bağlantılarını seçerken ırk, cinsiyet, yaş toplumsal statü, eğitim durumu, gelir, ulus, dil, din benzeri özelliklerinin ortak olduğu kişileri tercih etmektedir. *Seçicilik* terimi, ekoloji ve epidemiyoloji alanlarında sıkça kullanılır. Homofili ve seçicilik, benzer özelliklere sahip düğümler arasında bağlantıların yoğunlaştığını ifade etmek için kullanılır. Öte yandan modülerlik, homofili ve seçicilik nedeniyle oluşan sonuçları işaret eder. Modülerlik, ağda oluşan gruplaşmayı ölçmek için kullanılan bir ölçüdür. Ancak, matematiksel olarak üç ölçü de aynı şekilde hesaplanmaktadır. Çalışmanın devamında, her üç terim de *seçicilik* ile ifade edilecektir.

Seçiciliğin ölçülmesi için, bir seçicilik katsayısından yararlanılmaktadır. Seçicilik katsayısı, -1 ile +1 arasında değişmektedir. Eğer ağdaki bağlantılar tamamen benzer düğümler arasında gerçekleşiyorsa kusursuz bir benzerlik vardır ve katsayı 1’e eşittir. Eğer zıt düğümler arasında bağlantı varsa (kadın-erkek gibi) katsayı -1 olacaktır. Son olarak, bağlantıların seçimi tamamen tesadüfi ise katsayı sıfır olur.

Bu katsayı basit bir ifade ile düğümler arasındaki bağlantıların, beklenen bağlantılara oranı ile ölçülebilir. Seçicilik katsayısı, numerik ve skalar değerler için farklı hesaplanmaktadır. Skalar değerler için kovaryans formülünden yararlanılır. Numerik değerler için ise aynı türdeki düğümler arasındaki bağlantıların sayısı hesaplanır. Ayrıca tüm düğümlere ait toplam beklenen bağlantı sayısı hesaplanır. Bu iki hesaplamanın farkı, numerik değerler için seçicilik katsayısını göstermektedir.

$$\frac{1}{2} \sum_{ij} A_{ij} \delta(c_i, c_j) - \frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta(c_i, c_j)$$

Seçicilik sadece cinsiyet, ırk, eğitim durumu benzerlikleri için değil, aynı zamanda derece dağılımları konusunda da geçerlidir. Çok sayıda bağlantısı olan düğümler diğer düğümler tarafından daha çok tercih edilmektedir. Bu sınıfta popüleritesi fazla olan çocuklarla herkesin arkadaş olmak istemesi bu duruma verilebilecek ilk akla gelen örnektir. Düğüm derecelerine göre seçiciliğin hesaplanmasında da kovaryans hesaplamasından yararlanılmaktadır. [27, p. 171]

Seçicilik, enformasyonun yayılması üzerinde de çok etkilidir. Sosyal medyada bilginin yayılması homofili özelliği gösteren ağlarda daha hızlı gerçekleşir. [40] Buna karşılık, benzerlik ne kadar azsa homofili o kadar düşük olur. Bu durumda, ağı oluşturan düğümler heterojendir. Negatif-seçicilik adı verilen bu durum da, istenmeyen bir sonuç değildir, çünkü yeni fikirlerin farklı gruplar arasında paylaşılmasını kolaylaştırmaktadır.

Bazı kaynaklara göre kümeleme katsayısı ile homofili arasında ilişki bulunmaktadır.[21] Homofili arttığında, kümeleme de artış gösterir. Gruplar arasında belirgin farklılıklar olmaması, bir yandan var olan enformasyonun hızlı yayılmasını sağlıyor. Ancak, farklı gruplar arasında ortaya çıkan enformasyonun diğer gruplara yayılmasını önlemektedir. Zira, bu yayılma sadece gruplar arasındaki ince, zayıf bağlantılar üzerinden gerçekleşebilmektedir. Bu durum da, yeni fikirlerin, farklı gruplar arasında yayılmasını engelleyen bir bariyer oluşturmaktadır.

4. Sentetik (Üretilmiş) Ağlar

4.1. Sentetik Ağların Kullanım Amaçları

Sentetik ya da üretilmiş (*generated*) ağlar, önceden belirlenmiş bazı parametrelere göre tasarlanmış ağlardır. Düzenli, tesadüfi, ölçek bağımsız ağlar, küçük-dünya ağları ve hatta konumsal ağlar da belirli parametrelerde değişiklik yapılarak sentetik olarak üretilebilir. Bir algoritma yardımıyla çok sayıda düğüm ve bağlantı yapay bir şekilde üretilebilir ve belirli büyüklüklere kadar görselleştirilebilir. Bu süreçte Java veya benzeri bir nesne yönelimli programlama dilinden de yararlanmak mümkündür.[3] Son bir kaç yıldır, ağ analizi yazılımlarındaki gelişmelerle birlikte ortaya çıkan ticari ve ticari-olmayan pek çok yazılım sayesinde istenilen sayıda düğüm ve bağlantıya sahip, belli parametrelere göre tasarlanmış sentetik ağ üretmek mümkündür. Bu çalışmada da faydalanılmış olan bazı önemli yazılımlar hakkında istenirse ilgili kaynaklardan ayrıntılı bilgi edinilebilir. [41]–[45]

Üretilmiş ağlar, pek çok kaynakta tesadüfi, küçük dünya ve ölçek bağımsız olmak üzere üç, bazı kaynaklarda da düzenli ağlar da ilave edilerek dört temel model olarak incelenmektedir.[46] Şekil 4'te dört farklı ağ türüne ilişkin örnek gösterilmiştir. Bu örnekler ORA programında, 100 düğüm adedi için oluşturulmuş, sentetik örneklerdir. [47]

Üretilmiş ağların amacı iki maddede özetlenebilir:

1. Ağ yapısının dinamiklerini anlamak için,
2. Gerçek dünyadaki ağ yapısının simülasyonu için.

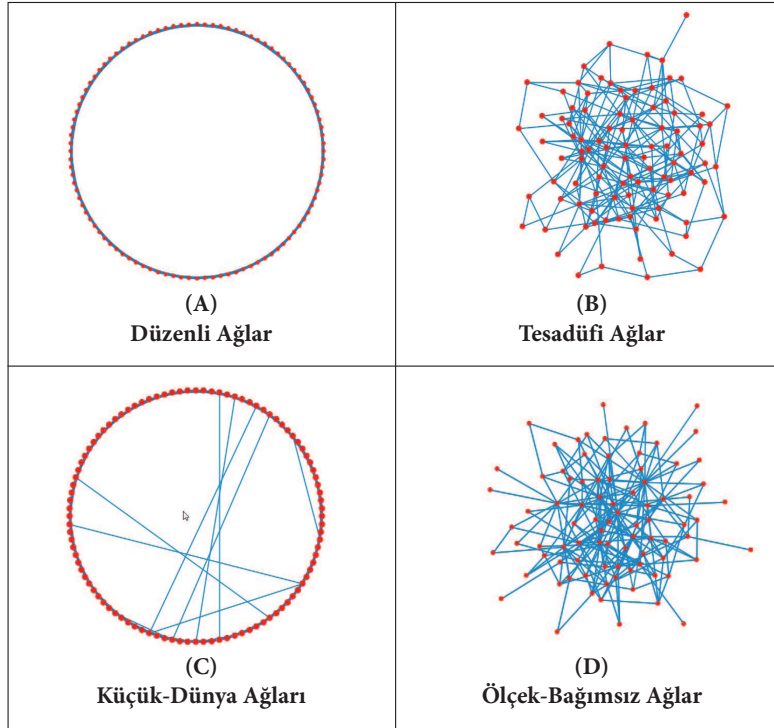
Düzenli ve tesadüfi ağlar, ağ yapısının dinamiklerini anlamak için kullanılırken, günümüzde küçük dünya ağları ve ölçek bağımsız ağlar gerçek dünyadaki ağların oluşturulmasında ve anlaşılmasında bir ölçüt (*benchmark*) olarak kullanılmaktadır.

4.2. Model Ağların Ortaya Çıkışı ve Çeşitleri

Üretilmiş sentetik ağ modelleri incelenirken bazı özellikleri ile birbirinden ayrıştıkları görülür. Gerçek hayatta karşılaşılan ağların da aynı özellikleri incelendiğinde hangi modele benzedikleri bu özelliklere göre anlaşılabilir. Derecelerin dağılımı, ortalama yol uzunluğu, kümeleme katsayısı ve merkezilik ölçüleri incelendiğinde, ağ modeli tanımlanabilir. [3]

Kronolojik bir şekilde gelişmiş olan dört önemli model ağ bulunmaktadır. Aslında model ağların gelişimi ağ biliminin gelişim tarihi ile paralellik arz etmektedir. Ağ modelleri, basit, düzenli ağlardan karmaşık sistemleri incelemeye uygun karmaşık modellere doğru bir ilerlemedir. Ortaya çıkış sırasına göre ağ modelleri şu şekildedir:

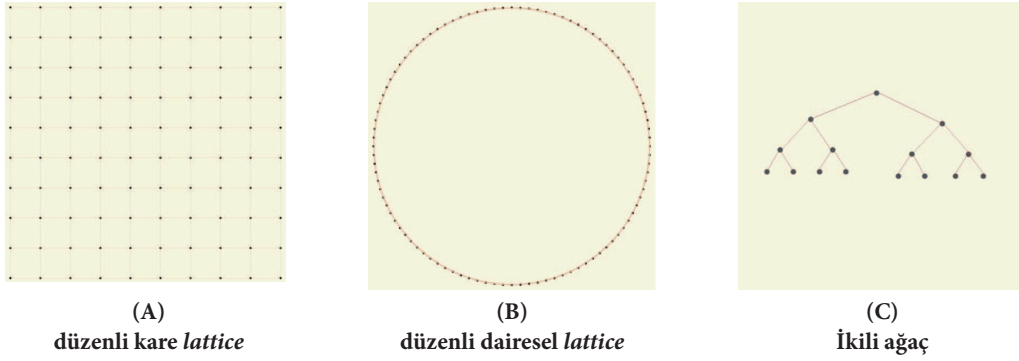
- i. Düzenli ağlar
- ii. Tesadüfi ağlar
- iii. Küçük-Dünya Ağları
- iv. Ölçek Bağımsız Ağlar



Şekil 4: Üretilmiş ağların türleri

Şekil 5'te üç farklı düzenli ağ örneği gösterilmektedir. Sadece kare şeklinde değil, daire, yıldız hatta hiyerarşik yani ağaç şeklinde de olabilir. (Şekil 4.2 A,B,C,)

Düzenli ağların başlıca özellikleri arasında; seyrek (*sparse*) ve bağlantılı olmaları (*connected*), ağ çapının küçük, ağın merkezine olan uzaklığının (*small central node radius*) ve ortalama yol uzunluğunun kısa (*small average shortest path*) olması sayılabilir.[3] Ayrıca düzenli ağlarda komşu sayıları eşit olduğundan derece dağılımı tekdüze (*uniform*) dağılıma uymaktadır.



Şekil 5: Düzenli ağ örnek şekiller

Düzenli ağların tersine tesadüfi ağlar belli bir oranda tesadüflük içermektedir ve bu nedenle entropi değeri sıfıra eşit değildir. Tesadüfi ağlar, olasılık teorisi ile ağ teorisinin kesişim kümesinde yer alırlar. Düğümler birbirleri ile p olarak ifade edilen bir olasılık değeri ile bağlanırlar. p değeri 0 ile 1 arasındadır. Tesadüfi ağ teorisi tek bir ağı incelemeyi, önceden belirlenmiş bir istatistiksel özelliği taşıyan olası bir ağ grubunun hepsinin özelliklerini inceler.[48, p. 9] Bu nedenle, belirli bir p değerine sahip bütün ağ örneklerinden oluşan bir ağ kümesi düşünülmelidir.

Gerçek dünyada ağlar, ne tam anlamıyla düzenli ne de tam anlamıyla tesadüfidir; fakat her ikisinin de önemli özelliklerini taşımaktadır. [2] Hiç bir biyolojik yapıda, bilgisayar ağında ya da sosyal ağda bu derece rastlantısal bağlantıların ortaya çıkması mümkün değildir. Ancak daha öncede belirtildiği gibi bu ağlar gerçek dünyadaki ağların yapısını anlamak için birer yardımcı araç niteliğindedir.

Bir tesadüfi ağda, herhangi iki düğüm arasında bir bağlantının olma olasılığı (p) olsun. Bu durumda (N) toplam düğüm sayısı ise düğüm başına ortalama bağlantı sayısı ($\bar{n}$) aşağıdaki şekilde hesaplanabilir. [49]

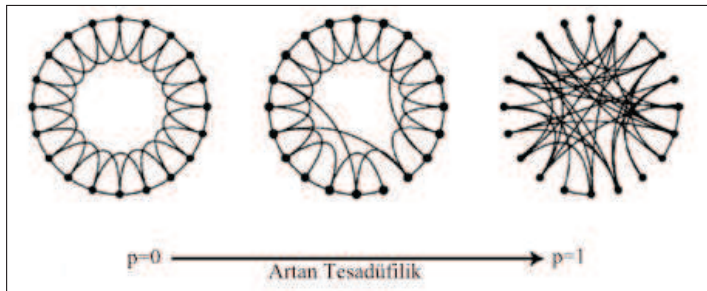
$$\bar{n} = p(N - 1)$$

Klasik tesadüfi ağların en önemli özelliklerinden birisi de *poisson* ya da *binom* dağılıma uygunluk gösteren derece dağılımıdır.

Bir diğer ağ modeli olan olan, küçük dünya ağları, *küçük dünya fenomeni* olarak bilinen durumdan ileri gelmektedir. Bu fenomen, bir ağda tesadüfi olarak seçilen herhangi iki düğümün, tanıdıklardan oluşan kısa bir zincirle birbirine bağlanması şeklinde açıklanabilir.

Milgram'ın küçük dünya deneyine dayanan bu ağ modeli *Watts-Strogatz* tarafından 1998 yılında oluşturulmuştur. O zamandan beri pek çok bilim insanı tarafından (*Barabási-Albert, Watts, Newman vs.*) çok sayıda varyasyonu ortaya konmuştur. [50]

Sosyal ağlar, küçük dünya ağlarının bazı özelliklerini taşımaktadır. Watts-Strogatz küçük dünya ağlarını, rastgele seçilen düğümleri arasında, yeni bağlantılar oluşmuş olan düzenli ağlar olarak tanımlamaktadır.[51] Şekil 6'da gösterilmiş olan Watts'ın küçük dünya ağları ile ilgili ünlü şekli de bu tanıma uymaktadır.[2] Şekilde görüldüğü gibi, $p=0$ olduğunda ağ tamamen düzenli, $p=1$ olduğunda tamamen tesadüfi olmaktadır.



Şekil 6: Düzenli ağların tesadüflüğünün artması [2, p. 245]

Bu değerin yani p değerinin düşük fakat sıfıra eşit olmadığı durumlarda ise, Şekil 4.3'de görülen pek çok lokal bağlantının, az sayıda da uzak bağlantının bulunduğu; ortalama yol uzunluğunun kısa, kümeleme katsayısının yüksek olduğu küçük-dünya ağ ortaya çıkmaktadır.[16]

En yeni ağ modeli olan ölçek bağımsız ağlar, şimdiye kadar üretilen ağlar içerisinde gerçek ağlarla en fazla benzerlik gösteren ağ modelidir.

Bilindiği gibi, gerçek hayatta ağlar durağan değildir. Yeni düğüm ve bağlantıların eklenmesi ya da bazı düğüm ve bağlantıların yok olması dolayısı ile sürekli değişmektedir. Ağın genişlemesi, ölçek bağımsız özelliklerin ilk akla gelen özelliğidir. İkincisi ise, "tercihli ya da imtiyazlı bağlantı" (*preferential attachment*) olarak tanımlanan özelliktir. Tercihli bağlantı, özelliğine, önceki bölümde seçici bağlantılar anlatılırken (*Bölüm 3.1.4*) kısaca değinilmiştir. Kısaca, tercihli bağlantı özelliği, seçicilik özelliğinin bir sonucu olarak tanımlanabilir.

Ölçek bağımsız model bu iki önemli özelliği sayesinde tesadüfi ağlardan ayrılmakta ve gerçek ağları temsil etmeye yaklaşmaktadır. [50]

4.3. Model Ağların Yapısal Özelliklerinin Karşılaştırılması

Üçüncü bölümde gerçek ağların yapısal özellikleri arasında dört önemli başlığa yer verilmiştir. Model ağların, yapısal özellikleri de birbirinden farklıdır. Bu yapısal özelliklerden bazıları, gerçek dünya örneklerine çok yakındır. Bazıları ise gerçek hayatta bu şekilde özellikler göstermeyen değerlere sahiptir.

Bu durum büyük ölçüde ağ modellerinin kronolojik gelişiminden kaynaklanmaktadır. Düzenli ağlar ve tesadüfi ağlar ilk ortaya çıkan ağ modelleri oldukları için daha çok matematiksel

hesaplamalara dayanmaktadır. Ancak bazı karmaşık düğüm bağlantı ilişkilerinin fark edilmesi zaman içinde gerçekleşmiştir. Bu nedenle daha sonra ortaya çıkan modellerde, öncüllerinin bu eksiklerini gidermek için yeni yöntemler ve parametreler ortaya konmuştur. Yapısal özelliklerdeki değişiklikler zaman içerisinde fark edilen bu değişikliklere göre gerçekleşmektedir.

Aşağıdaki bölümlerde, hem model ağları birbirinden ayırmayı, hem de gerçek dünya ağları ile karşılaştırmayı sağlayan önemli bazı öznelitliklere yer verilmiştir. Yapısal özelliklerin tanımları Bölüm 3'de ayrıntılı olarak yapıldığı için burada sadece gerekli görülen açıklamalara yer verilecektir.

4.3.1. Düğüm Derece Dağılımı

Düğüm derece dağılımının tanımından önceki bölümde bahsedilmişti. Ayrıca, derece dağılımlarının uygunluk gösterdiği önemli dağılımlar da kısaca açıklandı. Tekdüze, *Poisson*, *Binom*, üstel (*exponential*) ve güç yasası dağılımları, ağlar incelerken en çok karşılaşılan dağılımlar arasındadır.

Düzenli ağlarda düğüm derecesi her bir düğüm için aynı olduğundan düğüm derece dağılımı tekdüze dağılıma uygundur. Ancak, diğer üç ağ modelinde dağılımlar farklılık göstermektedir.

Tesadüfi ağlarda, derece dağılımı *binom* (n küçük ise) ya da *poisson* (n büyük ise) dağılımına uymaktadır. Yönsüz bir tesadüfi ağ $G_{n,p}$ şeklinde gösterilmektedir. Aşağıdaki eşitliklerde, tesadüfi ağlarda her iki dağılımın hesaplanışı karşılaştırılmıştır.

$$\begin{aligned} \binom{n-1}{k} p^k (1-p)^{n-1-k} & \quad (\text{Binom Dağılımı}) \\ \frac{(np)^k}{k!} e^{-np} & \quad (\text{Poisson Dağılımı}) \end{aligned}$$

Aşağıdaki eşitlik, tesadüfi ağlarda *poisson* dağılıma uygun olan düğüm derece dağılımının hesaplanışını göstermektedir.

$$P(k_i) = \binom{n-1}{k_i} p^{k_i} (1-p)^{n-1-k_i}$$

Yine de *poisson* dağılım gerçek hayattaki tüm ağları tanımlamada yetersiz kalmaktadır. Bu nedenle üstel dağılımın olduğu bir tür tesadüfi ağ da geliştirilmiştir. [14, p. 274] Ancak bu da tam olarak sorunun çözümü değildir, zira WWW, Internet benzeri pek çok deneysel çalışmada dağılımın bu üç dağılıma da uymadığı gözlenmiştir.

Üçüncü ağ modeli olan küçük dünya ağları, düğüm derece dağılımı açısından karşılaştırıldığında, diğer sentetik ağlardan farklılık göstermektedir. Tesadüflüğün derecesini gösteren, 0 ve 1 arasında değişen, p değeri arttıkça derece dağılımının özellikleri de değişmektedir. (Şekil 4.3) Yapılan çeşitli çalışmaların ortak özelliği, ağ genişledikçe herhangi iki düğüm arasındaki uzaklığı da logaritmik olarak artmasıdır. [50]

Küçük dünya ağları, diğer yapısal özellikleri açısından gerçek ağlara benzerlik göstermesine rağmen düğüm derece dağılımı açısından karşılaştırıldığında onlardan farklılaşmaktadır.

Bölüm 3.1.1’de belirtildiği gibi gerçek ağlarda derece dağılımı genellikle güç yasası dağılımına uygundur. Bu tür bir dağılım aynı zamanda ölçek bağımsız ağ modelleri için de geçerlidir. Hatta bu nedenle, güç yasası dağılımı yerine “ölçek-bağımsız dağılım” da denilmektedir. Ölçek bağımsız ağlarda, düğüm dereceleri aşağıdaki şekilde hesaplanmaktadır. k_i ortalama derece dağılımı göstermektedir. $P(k_i)$ ise tesadüfen seçilen her hangi bir düğümün k_i ortalama derece dağılımı olma olasılığını gösterir.

$$p(k_i) \sim k_i^{-\gamma}$$

Ölçek bağımsız ağları, tesadüfi ağlardan ayıran en önemli özelliklerden birisi ortalama derece dağılımı aynı olsa bile dağılımın şeklidir. Bir ağın ölçek bağımsız olup olmadığını anlamak için iyi yöntemi derece dağılımını incelemektir. Bunun için logaritmik ölçek (*log-log plot*) ve histogramdan yararlanılmaktadır. (Bknz. Bölüm 3.2.1)

Ölçek bağımsız ağ modelleri gerçek ağların incelenmesinde kullanılmaktadır ve günümüzde pek çok gerçek ağ –bütün gerçek ağlar bu dağılıma uymasa bile- bu dağılıma uymaktadır. Örneğin, birlikte yayın yapan yazarlardan oluşan bir ağ (*co-authorship network*) daha çok üstel dağılıma uymaktadır. Bir diğer örnek ABD’nin batısındaki elektrik şebekesinin yapısının güç yasası yerine üstel dağılıma daha uygun oluşudur. [14]

Güç yasası, tüm gerçek ağlar için geçerli değilse de, ezici bir çoğunluk için doğrudur. Bu durumda düğümlerin çoğu ortalama düğüm derecesinden daha az sayıda düğüme sahipken, pek azı ortalamanın üzerinde düğüm sayısına sahiptir. [2] Göbek (*hub*) adı verilen bu az sayıdaki düğüm enformasyonun dağılmasında çok önemli rol oynamaktadır. Bu göbekler sayesinde enformasyon hızla yayılmaktadır ancak öte yandan bu yapı da, ağı saldırılara karşı açık hale getirmektedir. Yani bir yanı ile hata toleransı düşerken, diğer taraftan saldırılara karşı korunmasız hale gelmektedir. [52, p. 10] Örneğin bir bilgisayar ağında göbekler zarar gördüğünde tüm ağ çökebilir. Enformasyon açısından değerlendirildiğinde ise göbeklere iletilen yanlış bilginin hızla ağın geri kalanına ulaşması riskidir.

4.3.2. Ortalama Yol Uzunluğu

Ortalama yol uzunluğu, tüm ağın geometrik görünümü konusunda çok önemli temel bilgiler vermektedir. Modern ağ tasarımı açısından bakıldığında ağın topolojik yapısı, enformasyonun farklı sistemler arasında dağılmasında ve söylentilerin ya da hastalıkların yayılmasında büyük rol oynamaktadır. [19] Bütün bu süreçlerde, eğer herhangi iki düğüm arasındaki uzaklık kısa ise dağılıma daha hızlı gerçekleşmektedir.

Tesadüfi ağlarda, düğüm derece dağılımının özellikleri her ne kadar gerçek ağların düğüm dereceleri ile uyum sağlamıyorsa da; ortalama yol uzunluğunda durum farklıdır. Ağdaki düğüm sayısı arttıkça ortalama yol uzunluğu logaritmik ölçekte değişim gösterecektir. Bunun sebebi ortalama yol uzunluğu (L) düğüm sayısı (N) arasındaki logaritmik ilişkidir. [48]

$$L \sim \ln N$$

Ortalama düğüm sayısının ölçek bağımsız ağlara yaklaştığı bazı özel durumlarda tesadüfi ağlarda düğümler arasındaki ortalama yol uzunluğu kısalmır.

$$p(k) \sim k^{-\alpha}$$

Bu normal dışı durum, ancak yukarıdaki formülde α değerinin $2 < \alpha < 3$ arasında değiştiği özel durumlarda gözlenir. [31]

Oysa, küçük-dünya özellikleri taşıyan bir ağda ortalama yol uzunluğu çok daha kısadır. Ortalama yol uzunluğu küçük dünya ağlarının iki önemli özelliğinden birisidir. (Diğer önemli özellik kümeleme katsayısıdır.)

Bu durum küçük-dünya fenomeninden ileri gelmektedir. Küçük p değerleri için tesadüfi olarak oluşan az sayıda bağlantı (zayıf bağlantılar) sayesinde, ağdaki düğüm sayısı fazla bile olsa ortalama yol uzunluğu kısalmaktadır.

Enformasyonun yayılması açısından değerlendirildiğinde, küçük-dünya ağlarında yayılma hem daha hızlı gerçekleşmektedir hem de bu özellikler sayesinde aynı enformasyon ağına büyük bir bölümüne yayılmaktadır. Örneğin, bir kişi gün içerisinde fiziki olarak farklı mekanlara gidip buradaki kişilerle iletişime geçebilir. Farklı gruplarla temasa geçebilir. Bu durum gruplar arasında kısa bağlantılar yaratacaktır. Kısa ortalama yol uzunluğu ağına farklı noktalarına aynı enformasyonun iletilmesini kolaylaştıracaktır. Bu özellik düzenli ve tesadüfi ağlarda, yani p değerinin bir ya da sıfır olması durumlarında gözlenemez. [53]

Yayılma hızının ölçüsü ortalama yol uzunluğudur. Başka bir deyişle yayılmanın hızı, ortalama yol uzunluğunun kısa olup olmamasına bağlıdır.

Ölçek bağımsız ağlarda ortalama yol uzunluğu benzer düğüm ve bağlantı sayısına sahip gerçek ağlardan ve tesadüfi ağlardan çok daha azdır. Bu sayı tesadüfi ağlarda anlatılan anomali durumunda gerçekleşen ortalama yol uzunluğu olan $\ln N$ değerine eşittir.

$$L \sim \ln N$$

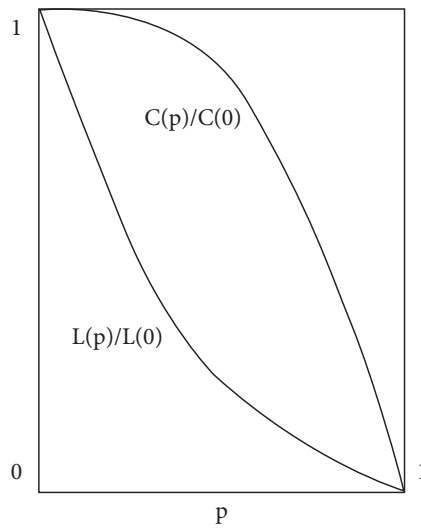
Ortalama yol uzunluğunun kısa olması yukarıdaki ölçek bağımsız ağlarda da, enformasyonun yayılması açısından çok elverişli bir zemin oluşturmaktadır. Ölçek bağımsız ağlarda, ortalama yol uzunluğunun kısa olması derece dağılımının özel olmasından kaynaklıdır. Derece dağılımı güç yasası dağılımına uygun olan, ölçek bağımsız ağlar, gerçek ağlara en yakın ortalama yol uzunluğu değerlerine sahiptir.

4.3.3. Kümeleme Katsayısı Büyüklüğü

Tesadüfi ağlarda, düğümlerin konumları önemli değildir. Bunun sebeplerinden biri her bir düğümün önceden belirlenmiş bir olasılık değeri (p) ile komşularının sayısının belirlenmiş olmasıdır. Bu homojenlik sebebi ile düğümlerin kümelenmesi azdır. Kümeleme katsayısı diğer ağlarla karşılaştırıldığında düşüktür. (Keeling & Eames, 2005, p. 300)

Ortalama yol uzunluğunun yanında, küçük dünya ağlarında gözlenen bir diğer önemli özellik de yüksek düzeyde kümelenmenin olmasıdır. Dorogovtsev ve Mendes'e göre küçük dünya modeli yüksek düzeyde kümelenme özelliğini düzenli ağlardan, kısa ortalama yol uzunluğunu da tesadüfi ağlardan miras almıştır. [46, p. 83] Bölgesel bir yoğunluk ölçüsü olan kümeleme katsayısı ile global bir ölçü olan ortalama yol uzunluğunun p 'nin bir fonksiyonu olarak değişimi Watts tarafından oluşturulan Şekil 7'de gösterilmektedir [2]

Aslında, küçük dünya ağlarında enformasyonun hızlı yayılmasının tek sebebi, ortalama yol uzunluğunun kısa olması değildir. Kümeleme katsayısının yüksek olması, ortalama yol uzunluğunun kısa olmasına sebep olmaktadır. Bu iki önemli özellik, düzenli ağların ve tesadüfi ağların, gerçek ağları inceleme noktasındaki yetersizliklerini ortadan kaldırmaktadır. Yayılma modelleri açısından bakıldığında, yüksek kümeleme katsayısı enformasyonun yayılma yeteneğini; kısa ortalama yol uzunluğu ise enformasyonun bulunma maliyetinin düşüklüğünü göstermektedir.[54] Ancak yine de düğümler arasında yüksek derecede bağlantınlık (*connectivity*) ve seçicilik (*assortativity*) bulundurmayan bu ağlar gerçek ağları incelemede yetersiz kalmaktadır.



Şekil 7: Ortalama yol uzunluğu ve kümeleme katsayısının p değerine göre değişimi [2, p. 245]

Ölçek bağımsız ağlar ve küçük dünya ağları karşılaştırıldığında kümeleme katsayısı açısından büyük bir eksiklerinin olduğu göze çarpmaktadır. Her ne kadar düğüm derece dağılımı gerçek ağlara benzer özellikler gösterse de; kümeleme katsayısı küçük dünya ağlarına kıyasla oldukça düşüktür. Ayrıca bu ağ modellerinin düğümleri arasında ilişki olmaması yani seçiciliğin tam anlamıyla görülmemesi de (sadece düğüm bağlantı sayısına göre bazı durumlarda), bu ağları gerçek ağlardan uzaklaştırmaktadır. Bu durumda bu zayıflıkları karşılayacak yeni modellere ihtiyaç duyulmaktadır. [55], [56]

4.3.4. Seçici Bağlantı

Seçicilik, gerçek ağlar ve sentetik ağları birbirinden ayırmak için kullanılabilecek en çarpıcı özelliklerden birisidir. Seçicilik, düzenli, tesadüfi ağlar için bir parametre değildir. Bunun doğal sonucu olarak küçük-dünya ağlarında da böyle bir özellik görülmez.

Newman'ın 2002 yılında yapmış olduğu çalışmada, gerçek ağlar ve bazı sentetik ağ modelleri seçicilik katsayıları hesaplanarak, karşılaştırılmıştır.[57, p. 2] Şekil 8'de görüldüğü gibi gerçek ağ-

larda, düğümler arasında pozitif ya da negatif seçicilik görülmektedir. (Fizik, biyoloji matematik yazarları ya da aktörler arasındaki dayanışma ya da şirket yöneticileri benzeri gerçek ağlarda pozitif-seçicilik; Internet, WWW, proteinlerin kimyasal etkileşimi, sinir ağları ve yemek ağları, teknolojik ağlarda bazı gerçek ağ örneklerinde negatif-seçicilik gözlenmektedir.)

	Ağlar	n	r
Gerçek Ağlar	Fizik alanında ortak yazarlar	52 909	0.363
	Biyolojik alanında ortak yazarlar	1 520 251	0.127
	Matematik alanında ortak yazarlar	253 339	0.12
	Film-aktör ağlar	449 913	0.208
	Şirket yöneticileri	7 673	0.276
	İnternet	10 697	-0.189
	WWW (World Wide Web)	269 504	-0.065
	Protein etkileşimleri	2 115	-0.156
	Sinir ağları	307	-0.163
	Beslenme ağları	92	-0.176
Modeller	Tesadüfi ağlar		0
	Callaway vd.		$\delta/(1+2\delta)$
	Barabási ve Albert		0

Şekil 8: Farklı ağların büyüklüğüne (n) ve seçicilik katsayısına (r) göre karşılaştırılması [57, p. 2]

Şekilde görüldüğü gibi, neredeyse tüm sentetik ağ örneklerinde seçicilik katsayısı sıfıra yakındır. Sadece ölçek bağımsız ağlarda, tercihli bağlantı özelliğinin bir sonucu olarak düğüm sayısı fazla olan düğümler, diğer düğümler tarafından daha çok tercih edilmektedir.

Bölüm 3.1.4’de vurgulandığı gibi sadece düğüm derece dağılımları değil, pek çok farklı öz-nitelikten dolayı (dil, ırk, cinsiyet vb.) düğümler arasında seçicilik ortaya çıkabilir. Bu haliyle ölçek bağımsız ağlardaki tercihli bağlantı özelliği seçiciliği gerçek hayatta rastlandığı boyutları ile modellemekten uzak kalmaktadır.

5. Sonuç ve Değerlendirme

Her dört özellik de sadece gerçek ağların anlaşılması açısından değil, aynı zamanda sentetik ağların yorumlanması için de gereklidir. Hem gerçek ağlarla, hem de üretilmiş ağlarla yapılmış olan pek çok deneysel çalışmada, ağ türlerine göre bir takım sonuçlar elde edilmiştir. Mitchell tarafından 2006 yılında yapılmış çalışmada gerçek ve sentetik ağlar derece dağılımı, ortalama yol uzunluğu ve kümeleme katsayısına göre karşılaştırılmıştır.[16, p. 6] Mitchell tarafından hazırlanan bu tablo seçici bağlantılar eklenerek ve yeniden yorumlanarak Tablo 1’de yeniden oluşturulmuştur.

Tablo 1: Gerçek ağlar ve üretilmiş ağların karşılaştırılması

Ağ modeli	Düğüm Derece Dağılımı	Ortalama Yol Uzunluğu	Kümeleme Katsayısı	Seçici Bağlantı
Gerçek Ağlar	güç yasası (genellikle)	yüksek	düşük	pozitif yada negatif seçicilik
Düzenli Ağlar	sabit	yüksek	yüksek	yok
Tesadüfi Ağlar	poisson-binom	düşük	düşük	yok
Küçük Dünya Ağları	p 'ye bağlı	yüksek	düşük	yok
Ölçek-Bağımsız Ağlar	güç yasası	yüksek	düşük	kısmen

Tablo 1, önceki bölümlerde yapılan değerlendirmelerin bir özeti niteliğindedir. Mitchell'in 2006 yılındaki çalışmasındaki tablodan esinlenerek oluşturulmuş olan Tablo 1 gerçek ağlar ve sentetik ağların önemli bulunan özellikleri karşılaştırılarak elde edilmiştir.[16] Ağlar büyük ve karmaşık yapılar oldukları için çoğu özellikleri özetleme ve genellemeler yapmaya müsait değildir. Uzun yıllar süren çalışmalar sonucunda ortaya konmuş bu dört karakteristik özellik için bile ağ modellerine göre istisnalar ortaya çıkmaktadır.

Karmaşık ağ yapılarında saklı bulunan, keşfedilecek daha başka pek çok ayırıcı özellik mevcuttur. Buradaki temel problem, analiz yapılırken aşırı derecede basit modeller oluşturarak problemleri aşırı basite indirgemekten kaçınmaktır. Diğer yandan, kabul edilebilir düzeyde karmaşıklığı koruyarak kurulan modeller ve seçilen parametrelerin doğruluğunun test edilmesi ve ispatlanması da çok güçtür. Zira bu kadar büyük karmaşık mekanizmalar için yapılan yorumların sadece o özgün duruma uygun olması mümkündür. Başlangıç koşullarındaki küçük bir değişime dahi son derece duyarlı olan bu karmaşık yapılarda genel geçer ilkeler ve özellikler tespit etmek çok güçtür.

Yararlanılan Kaynaklar

- [1] A.-L. Barabási, *Linked: The new science of networks*, vol. 71. Perseus Publishing, 2002, p. 123.
- [2] D. J. Watts, *The "New" Science of Networks*, vol. 30, no. 1. 2004, pp. 243–270.
- [3] T. G. Lewis, *Network Science*. 2009.
- [4] Infographics, "Statistics for Social Media Sites," 2012. [Online]. Available: <http://cdn2.business2community.com/wp-content/uploads/2012/12/Social-Growth-infographic4.png>. [Accessed: 05-Feb-2013].
- [5] Central Intelligence Agency, "CIA - The World Factbook," 2012. [Online]. Available: <https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html>. [Accessed: 16-Feb-2013].
- [6] D. Bennett, "The Dunbar Number, From the Guru of Social Networks," *Bloomberg Businessweek Technology*, pp. 1–6, 10-Jan-2013.
- [7] V. Mahajan and R. Peterson, *Models for innovation diffusion*. 1985.
- [8] T. Valente, "Social network thresholds in the diffusion of innovations," *Soc. Networks*, vol. 18, no. 1, pp. 69–89, Jan. 1996.
- [9] J. Cointet and C. Roth, "Information diffusion in realistic networks," *Main*, pp. 1–4, 2007.
- [10] E. M. E. Rogers, *Diffusion of innovations*. Free Pr, 1995.
- [11] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge Univ Pr, 2010.
- [12] R. Albert and A.-L. Barabási, "Statistical Mechanics of Complex Networks," vol. 74, no. January, 2002.
- [13] B. Schölkopf, "Structure and Dynamics of Information Pathways in Online Media," 2013.
- [14] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–76, Mar. 2001.
- [15] M. E. J. Newman, "Random graphs as models of networks," *Handb. graphs networks*, no. 1, 2003.
- [16] M. Mitchell, "Complex systems: Network thinking," *Artif. Intell.*, vol. 170, no. 18, pp. 1194–1212, 2006.
- [17] C. Gros, *Complex and adaptive dynamical systems*, vol. 1. 2008, p. 2865.
- [18] S. N. Dorogovtsev, "Evolution of networks," 2001.
- [19] A. Fronczak, P. Fronczak, and J. A. Holyst, "Average path length in random networks," no. 3, pp. 1–4, 2008.
- [20] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, Jan. 2003.
- [21] M. O. Jackson, *Social and Economic Networks*. 2008.
- [22] D. Radev and D. L. C. Mack, "Power Law Degree Distributions," pp. 1–4, 2008.
- [23] A.-L. Barabási, R. Albert, and H. Jeong, "Diameter of the World-Wide Web," *Nature*, vol. 401, no. September, pp. 398–399, 1999.
- [24] B. Huberman and L. Adamic, "Growth dynamics of the World-Wide Web," *Nature*, vol. 401, p. 131, 1999.
- [25] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science (80-.)*, vol. 286, no. October, pp. 509–512, 1999.
- [26] D. Hutchison and J. C. Mitchell, *Network Analysis*. 2005.
- [27] M. E. J. Newman, *Networks: An Introduction*. 2010.
- [28] Network-science.org, "From sparsely to densely connected networks." [Online]. Available: <http://www.network-science.org/highly-connected-society-dense-social-complex-networks.html>. [Accessed: 19-Apr-2013].
- [29] S. Milgram, "The small world problem," *Psychol. Today*, vol. 1, no. 1, pp. 61–67, 1967.

- [30] C. A. R. Pinheiro, *Social network analysis in telecommunications*. John Wiley & Sons, Inc., 2011, p. 95.
- [31] A. Abraham, A.-E. Hassanien, and V. Snasel, *Computational Social Network Analysis*. 2010.
- [32] K. M. Carley, J. Reminga, J. Storrick, and D. Columbus, "ORA User 's Guide 2010," 2010.
- [33] S. Bornholdt, H. Schuster, and J. Wiley, *Handbook of graphs and networks*. 2003.
- [34] SOCNET Archives, "Assortativity, Homophily, Modularity," 2009. [Online]. Available: <http://lists.ufl.edu/cgi-bin/wa?A2=ind0905&L=SOCNET&D=0&P=27695>. [Accessed: 25-Apr-2013].
- [35] M. E. J. Newman, "Difference between assortativity, homophily, modularity," 2009. [Online]. Available: <http://zvonka.fmf.uni-lj.si/netbook/doku.php?id=private:socnet>. [Accessed: 25-Apr-2013].
- [36] P. J. Carrington, J. P. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*. 2005, p. 345.
- [37] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '08*, p. 160, 2008.
- [38] S. Wasserman and K. Faust, *Social network analysis*. 1994.
- [39] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a Feather : Homophily in Social Networks Source," *Annu. Rev. Sociol.*, vol. 27, no. 2001, pp. 415–444, 2001.
- [40] B. Furht, *Handbook of Social Network Technologies and Applications*, vol. 129. Springer-Verlag New York Inc, 2010, p. 736.
- [41] K. M. Carley, J. Reminga, J. Storrick, and D. Columbus, *ORA User's Guide 2011*. 2011.
- [42] A. T. Inc., "UCINET 6.3.0.1." 2010.
- [43] V. Batagelj and A. Mrvar, "Pajek 2.02." 2010.
- [44] Sonomine, "Sonomine 2.0." 2010.
- [45] D. Hansen, B. Shneiderman, and M. Smith, *Analyzing social media networks with NodeXL*. 2010.
- [46] L. Hamill and N. Gilbert, "Simulating large social networks in agent-based models: A social circle model," *Emerg. Complex. Organ.*, vol. 12, no. 4, pp. 78–94, 2010.
- [47] K. M. Carley and C. Team, "ORA Netscenes v3.0.0.2." Carnegie-Mellon University, 2013.
- [48] S. Havlin and R. Cohen, *Complex Networks: Robustness and Function*. 2010.
- [49] M. J. Keeling and K. T. D. Eames, "Networks and epidemic models," *J. R. Soc. Interface*, vol. 2, no. 4, pp. 295–307, Sep. 2005.
- [50] C. Y. Huang, C. T. Sun, and H. C. Lin, "Influence of local information on social simulations in small-world network models," *JASSSTHE J. Artif. Soc. Soc. Simul.*, vol. 8, no. 4, pp. 1–26, 2005.
- [51] F. Alkemade and C. Castaldi, "Strategies for the Diffusion of Innovations on Social Networks," *Comput. Econ.*, vol. 25, no. 1–2, pp. 3–23, Feb. 2005.
- [52] X. F. Wang and G. Chen, "Complex Networks : Scale-Free and Beyond," vol. 3, no. 2, 2003.
- [53] C.-Y. Huang, "A Novel Small-World Model: Using Social Mirror Identities for Epidemic Simulations," *Simulation*, vol. 81, no. 10, pp. 671–699, Oct. 2005.
- [54] R. Kali, "Social Embeddedness and Economic Governance : A Small World Approach," no. January, 2003.
- [55] L. Hamill and N. Gilbert, "A simple but more realistic agent-based model of a social network," *Structure*, 2008.
- [56] Q. Yan, L. Wu, and L. Zheng, "Social network based microblog user behavior analysis," *Phys. A Stat. Mech. its Appl.*, vol. 392, no. 7, pp. 1712–1723, Apr. 2013.
- [57] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 2, no. 4, pp. 1–5, 2002.



Şebnem AKAL / sebnemakal@marmara.edu.tr

Şebnem Akal was born in 1978 in Tekirdağ. She completed her undergraduate education in Marmara University Business Administration Department in German language. She successfully graduated from Marmara University Master's Degree of Business Informatics in 2006 and Istanbul University PhD of Quantitative Methods in 2010. She has been working in Marmara University Department of Business Informatics since 2004. She works as a Lecturer Doctor in Marmara University Department of Business Informatics. She is interested in scientific areas including network analysis, graph theory, data mining and business intelligence applications.