

Farklı Nitelikteki Biyolojik Ağların Entegrasyonu ve Yerel Topolojik Özellik Vektörleri Tabanlı Karşılaştırılması

Program Kodu: 1001

Proje No: 114E111

Proje Yürütücüsü:
Doç. Dr. Tolga Can

Danışman:
Doç. Dr. Yeşim AYDIN-SON

Bursiyer:
Andaç AKARSU

ARALIK 2016
ANKARA

ÖNSÖZ

TÜBİTAK 1001 programı kapsamında 114E111 proje numarası ile 2 yıl boyunca desteklenen bu projede farklı biyolojik olayları belirten farklı büyük ölçekli biyolojik ağlar entegre edilmiş ve alt çizge örüntülerinin sayılması ile ağlar için topolojik özellik vektörleri oluşturulmuştur. Farklı doku ve hastalıklara ait ağların karşılaştırılması ile dokulara ve hastalıklara özgü alt çizge örüntüleri belirlenmiştir. Karşılaştırılan ağların hizalanıp benzer kısımlarının da etkili görselleştirilebilmesi için hizalanmış ağlar için özelleşmiş bir ağ yerleştirme (İng. graph layout) yöntemi geliştirilmiştir. Bir doktora ve bir yüksek lisans tezi kapsamında yapılan proje çalışmalarından bir konferans bildirisi ve bir dergi makalesi üretilmiştir.

İÇİNDEKİLER

Önsöz	i
İçindekiler	ii
Tablolar	iii
Şekiller	iv
Özet	v
Abstract	vi
Giriş	1
Literatür Özeti	2
Gereç ve Yöntem	5
Bulgular	10
Tartışma ve Sonuç	17
Kaynakça	18
Ekler Listesi	20

TABLULAR

Tablo 1. Pathway Commons'dan elde edilen ağıdaki etkileşim tipleri ve sayıları	5
Tablo 2. İki protein arasındaki farklı etkileşimlerin kodlanması	7

ŞEKİLLER

Şekil 1. Eşleşen alt-çizgelerin önerilen çizge yerleştirme yöntemi ile görselleştirilmesi	10
Şekil 2. 141 doku/hastalık ağında istatistiksel olarak en anlamlı çizgecik	11
Şekil 3. En fazla z-puanı varyansına sahip anlamlı çizgecik	12
Şekil 4. Yüksek z-puanı varyansına sahip bir başka anlamlı çizgecik	12
Şekil 5. 21 beyin dokusunun çizgecik imzaları kullanılarak hiyerarşik kümelenmesi	13
Şekil 6. Talamus lateral çekirdek dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik	15
Şekil 7. Prostat bezi dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik ...	15
Şekil 8. Deri dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik	16

ÖZET

Bu projede, entegre genom-ölçekli ağların yönlü çizgecik imzaları kullanılarak analiz edilmesi için yöntemler geliştirilmiştir. Ek olarak, hizalanmış ağların etkili görselleştirilebilmesi için bir ağ yerleştirme yöntemi önerilmiştir.

Büyük hacimli verilerin heterojenliği nedeniyle entegre genom-ölçekli ağların analizi zor bir araştırma problemidir. Biyolojik ağların karakterizasyonu için çizgecik sayıları gibi birçok topolojik ölçüt bulunmaktadır. Bu projede, etiketli ve çok yönlü çizgelerdeki küçük alt-çizge örüntülerini saymak için yöntemler önerilmiştir. Proje kapsamında analiz ettiğimiz ağları Pathway Commons veritabanından elde ettiğimiz ve *H. sapiens* genlerini kapsayan fiziksel etkileşimleri, gen düzenleyici etkileşimleri ve metabolik etkileşimleri içeren ağları entegre ederek elde ettik. Bu entegre ağı daha sonra büyük ölçekli bir transkriptom profillemeye çalışmasını kullanarak doku ve hastalık bazında filtreledik ve farklı doku ve hastalıklar için özel entegre ağlar elde etmiş olduk. Yönsüz protein-protein etkileşim ağları için önerilmiş olan çizgecik sayma fikrini geliştirerek çok yönlü ve etiketli ağlara uyguladık ve her bir ağ için çizgecik sayı vektörlerini oluşturduk. Bu çizgecik sayılarının istatistiksel anlamlandırmasını rastgele ağlardan elde ettiğimiz çizgecik sayıları ile karşılaştırarak gerçekleştirdik. Çizgecik sayılarının etkinliğini farklı doku ve hastalıklardaki çizgecik sayılarını karşılaştırarak ve farklı durumlardaki ağları kümeleyerek gösterdik. Elde ettiğimiz sonuçlar entegre biyolojik ağlarda çok sayıda istatistiksel anlamlı çizgeciğin bulunduğunu ve ağların kümelenmesi ve sistem seviyesinde doku ve hastalık ağlarının karşılaştırılması için çizgecik sayılarının etkili bir gösterim olduğunu ortaya koymaktadır. Ek olarak, önerilen çizge yerleştirme algoritmasının karşılaştırılan ağlardaki benzerlik ve farklılıkların görselleştirilmesi için etkin bir yöntem olduğu gösterilmiştir.

Anahtar kelimeler: entegre biyolojik ağlar, ağ karşılaştırma, yönlü çizgecikler, hizalanmış ağların görselleştirilmesi.

ABSTRACT

In this project, we developed a framework for the analysis of integrated genome-scale networks using directed graphlet signatures. In addition, we developed a novel graph layout algorithm specific for visualizing aligned networks.

Analysis of integrated genome-scale networks is a challenging problem due to heterogeneity of high-throughput data. There are several topological measures, such as graphlet counts, for characterization of biological networks. In this project, we present methods for counting small sub-graph patterns in integrated genome-scale networks which are modeled as labeled multidigraphs. We have obtained physical, regulatory, and metabolic interactions between *H. sapiens* proteins from the Pathway Commons database. The integrated network is filtered for tissue/disease specific proteins by using a large-scale human transcriptional profiling study, resulting in several tissue and disease specific sub-networks. We have applied and extended the idea of graphlet counting in undirected protein-protein interaction (PPI) networks to directed multi-labeled networks and represented each network as a vector of graphlet counts. Graphlet counts are assessed for statistical significance by comparison against a set of randomized networks. We present our results on analysis of differential graphlets between different conditions and on the utility of graphlet count vectors for clustering multiple condition specific networks. Our results show that there are numerous statistically significant graphlets in integrated biological networks and the graphlet signature vector can be used as an effective representation of a multi-labeled network for clustering and systems level analysis of tissue/disease specific networks. In addition, the proposed graph layout algorithm can be used to visualize the similarities and differences between aligned regions of these networks.

Keywords: integrated biological networks, network comparison, directed graphlets, visualization of network alignments.

1. GİRİŞ

Hücrel ağlar DNA, RNA, protein ve metabolitler arasındaki etkileşimleri göstermeleri açısından büyük önem taşırlar. Gen düzenleyici ve sinyal iletim ağları gen baskılama ve aktive etme olaylarının hücrede nasıl gerçekleştiğini ve bu işlemlerin hangi proteinlerin üretilmesine sebep olduğunu göstermektedir. Protein etkileşim ağları, protein bileşikleri oluşturmak ve proteinlerin başka proteinleri aktifleştirmesi gibi, proteinler arası etkileşimleri göstermektedir. Metabolik ağlar, enerji üretimi ve madde sentezleme gibi yaşam döngüsü için temel olan metabolik reaksiyonların birbirleriyle olan etkileşimlerini göstermektedir.

Büyük hacimli genomik, transcriptomik, metabolik ve proteomik verilerinin kamuya açık veritabanlarında günden güne artması ile heterojen ve dinamik biyolojik ağların entegre bir şekilde analizi mümkün hale gelmiştir. Pathway Commons (Cerami vd., 2011), BioGRID (Stark vd., 2006) ve Human Protein Reference Database (HPRD) (Keshava Prasad vd., 2009) gibi veritabanları, çeşitli büyük ölçekli ve küçük ölçekli deney sonuçlarından derlenen, genler, proteinler ve kimyasal bileşenler arasında ilişkileri toplu bir şekilde organize etmekte ve kullanıma hazır bir durumda araştırmacılara sunmaktadırlar. Ek olarak, BioPAX (Demir vd., 2010) gibi gösterimler ile farklı tür biyolojik ağların farklı veritabanları ve uygulamalar arasında taşınması ve kullanımı mümkün hale gelmektedir. Bu tür veritabanlarında her ne kadar farklı türde etkileşimler bulunsada gen düzenleyici ağlar, metabolik ağlar ve fiziksel protein-protein etkileşim ağlarının tek bir entegre ağda analizi hala bir araştırma problemi olarak durmaktadır.

Bu projede, etiketli ve çok yönlü çizgeler olarak modellenen entegre heterojen ağlar üzerinde olan bütün küçük alt-çizge örüntülerini saymak için yöntemler geliştirdik. Farklı dokular ve hastalıklara özgü 141 genom-ölçekli ağ elde ettik. Bu ağlarda hem fiziksel protein etkileşimleri, hem metabolik etkileşimler, hem de gen düzenleyici etkileşimler ağdaki kenarlar üzerinde farklı etiketler olarak gösterilerek entegre edildi. Daha önce yönsüz protein-protein etkileşim ağları için önerilmiş olan çizgecik sayma fikrini geliştirerek yönlü ve birden fazla etiketli kenarlar içerebilen ağlara uyarladık ve her bir ağ için çizgecik sayı vektörlerini oluşturduk. Sadece 2 ve 3 düğümlü alt çizgeler sayıldığında bile olası çizgecik türleri on binlere çıktığı için hem ağda en az geçme sayısını eşik değeri kullandık hem de istatistiksel olarak anlamlı olan çizgecikleri belirlemek için rastgele ağlardan elde ettiğimiz çizgecik sayılarını kullandık. Elde ettiğimiz çizgecik sayılarının biyolojik olarak anlamlandırılması ve birçok farklı analiz için faydalı olduğunu göstermek için doku ve hastalıklar arasında farklı sayılarda gözükken çizgecikleri tespit ettik ve ayrıca sadece çizgecik imzalarını kullanarak farklı durumlardaki ağları doğru şekilde kümeleyebildiğimizi gösterdik.

Proje önerisinde belirtmiş olduğumuz temel amaçlar olan 1) protein etkileşim, metabolik, gen düzenleyici ve sinyal iletim ağlarını birleştirerek sistem seviyesinde ve hastalık, zaman, doku vb. farklı durumlara bağlı etkileşim ağları oluşturmak ve 2) bu ağları karşılaştırmak için yerel topolojik özellik vektörleri tabanlı etkin bir yöntem geliştirmek amaçlarını proje sonunda gerçekleştirmiş bulunuyoruz. Bu hedeflere ulaşmak amacıyla tanımlanmış 8 iş paketi de sonuçlanmıştır. Proje çıktısı olarak ACM-BCB 2016 konferansı kapsamında yapılan CNB-MAC (İşlemsel Ağ Biyolojisi – Modelleme, Analiz ve Kontrol) çalıştayında bir bildiri sunulmuş ve bu bildirin uzun hali BMC Bioinformatics dergisinde yayınlanmak üzere kabul edilmiştir (bildiri ve makale Ekler listesinde sunulmaktadır). Final raporunun yazılması sırasında dergi makalesi üzerinde editör ofisinden gelen son şekilsel düzeltmeler yapılmış ve makale şu anda basım aşamasındadır. Arzu Burçak Sönmez doktora tezini ve Andaç Akarsu da yüksek lisans tezini hazırlamaktadır.

2. LİTERATÜR ÖZETİ

Son yıllarda teknolojik gelişmelerin de yardımıyla yüksek miktarda hücrel etkileşim verisi elde edilmiştir. Yüksek miktardaki bu veriler üzerinde yapılan çalışmalar araştırmacılara, hücrel faaliyetlerin ve hastalık mekanizmasının anlaşılması bakımından çok önemli fikirler vermektedir. Bu alanda yapılan çalışmalarda etkileşim ağlarının karşılaştırmalı analizi önemli bir yer tutmaktadır.

Büyük ölçekli etkileşim ağlarını yapısal olarak karşılaştırırken çizge eşbiçimlilik (İng. graph isomorphism) problemi karşımıza çıkar. Bu problemin çözümü özellikle büyük ölçekli etkileşim ağlarında uzun süreler aldığından farklı çözüm yolları aranmıştır. Çeşitli organizmaların protein etkileşim ağları üzerinde yapılan bir çalışmada (Przulj, 2007), bu probleme çözüm olarak çizgecik (İng. graphlet) adı verilen bölgesel yapılardan yararlanılmıştır. Çizgecikler, büyük çizgelerin, bağlantılı, eşbiçimli olmayan, türetilmiş (İng. induced) altçizgeleridir. Bu yapılar ağların yerel topolojik benzerliklerini ölçmek için kullanılmıştır. Karşılaştırılan etkileşim ağlarında, bu bölgesel yapıların tekrar etme sayıları saptanarak bir dağılım elde edilmekte, bu dağılımlar birbirleriyle karşılaştırılmaktadır. Önerilen yöntem yüksek sayıda karşılaştırma kriteri içerdiğinden, ağlar arası benzerlik yakalama olasılığı yüksektir.

Ancak, Przulj'in çalışmasında çizgecik adı verilen bölgesel yapıların tekrar etme sayıları rastgele modellenmiş bir etkileşim ağıyla karşılaştırılmamış böylece istatistiksel olarak

anamlı olup olmadıđı kontrol edilmemiřtir. Ayrıca karřılařtırma iřlemi sadece protein etkileřim ađları üzerinde yapılmıřtır. Üzerinde alıřılan etkileřim verilerinin yön bilgisi olmadıđından, izgecik adı verilen yapılarda da yön bilgisi dikkate alınmamıřtır.

Yapılan bir diđer alıřmada, bir etkileřim ađına ait motif adı verilen, istatistiksel olarak anlamlı ađ bađlantı örüntülerini saptamak için bir algoritma geliřtirilmiřtir (Milo vd., 2002). Bir örüntünün ađ motifi özelliđi gösterip göstermediđini belirlemek için, gerek etkileřim ađı üzerindeki tekrarlanma sayısı ile gerek etkileřim ađının yapısal özelliklerini taşıyan rastgele oluřturulmuř model ađ üzerindeki tekrarlanma sayısı karřılařtırılmıřtır. Bu uygulamada da etkileřim ađları izgelerle temsil edilmiřtir. Bu algoritmaya göre, öncelikle, farklı sayılarda düđüm noktasına sahip alt izgelerin, gerek etkileřim ađını sembolize eden izge üzerinde bulunma sayısı tespit edilerek, her n-düđüm noktasına sahip altizge için bu sayılar kaydedilmiřtir. Aynı iřlem rastgele oluřturulmuř model ađı temsil eden izge üzerinde tekrar edilmiřtir. Sonuç olarak, n-düđüm noktası ieren alt izge sayıları karřılařtırılmıřtır. Bu karřılařtırma sonucunda n-düđüm noktasına sahip bir altizgenin, rastgele oluřturulmuř ađda bulunma sıklıđının, gerek ađda bulunma sıklıđına eřit ya da ondan yüksek olma olasılıđı belli bir limit deđerin altındaysa, bu altizgeye, ađ motifi adı verilmiřtir. Bu uygulamayı ilk olarak transkripsiyon düzenleme ađları üzerinde gerekleřtirmiřlerdir. Bu ađlar yönlü ađlar oldukları için etkileřim yönleri de dikkate alınmıřtır (Milo vd., 2002). Yaptıkları bařka bir alıřmada ise (Yeger-Lotem vd., 2004), bu yöntem biraz daha geliřtirilmiř ve transkripsiyon düzenleme ađları ve protein etkileřim ađlarının birleřiminden elde edilen etkileřim ađı üzerinde uygulanmıřtır. İki farklı etkileřim ađı kullanıldıđı için, motifler belirlenirken, etkileřim yönlerinin yanında iki farklı etkileřim eřidi de dikkate alınmıřtır. Motiflerin tekrar etme sayıları rastgele ađlarla karřılařtırılıp belli bir limit deđere göre deđerlendirildiđi için izgeciklere göre istatistiksel olarak daha anlamlıdır ancak bu alıřma sadece iki eřit etkileřimi dikkate almıřtır.

Farklı organizmaların etkileřim ađlarını oluřturan moleküller arası etkileřim verileri birbirlerini tamamlayıcı niteliktedir. Örnek olarak, metabolik bir reaksiyonda görev alan bir enzim aynı zamanda protein etkileřim ađında adı geen bir protein ve gen düzenleyici bir etkileřimin elemanı olabilir. Sonuç olarak hücrenel etkileřimleri daha kapsamlı ve ok yönlü inceleyebilmek için farklı cinsten etkileřim ađlarının birleřtirilmesi önemlidir. Bu nedenle farklı etkileřim tiplerinin birlikte deđerlendirildiđi alıřmalar son dönemlerde önem kazanmıřtır. Örnek olarak, protein etkileřim ađları ve gen ifade verileri birlikte kullanılarak hastalıkların altında yatan iřleyiřin aydınlatıldıđı alıřmaları gösterebiliriz (Cline vd., 2007; Maraziotis vd., 2007).

Yapılan bir başka çalışma ise daha fazla etkileşim tipini birleştirmektedir (Kamburov vd., 2009). Çalışmanın güncel halinde (Kamburov vd., 2013), 30 farklı veritabanından alınan farklı türdeki etkileşim verileri ConsensusPathDB veritabanında toplanmıştır. Bu veritabanı, bir arayüz sayesinde birleştirilmiş etkileşim verilerinin sorgulanmasını ve görüntülenmesini sağlamaktadır. Bu veritabanında biyokimyasal reaksiyon, gen regülasyonu ve fiziksel etkileşim olmak üzere üç farklı etkileşim çeşidi dikkate alınmıştır. Ayrıca etkileşen moleküllerin özelliklerine göre bu etkileşim çeşitleri kendi içlerinde de birbirinden ayrılmaktadır. Bu etkileşim çeşitleri de 7 farklı kenar sembolüyle görsellenerek kullanıcının bilgisine sunulmuştur.

Paxtools yazılımlar grubu (Demir vd., 2013) entegre BioPAX ağlarını sorgulamak, görselleştirmek ve farklı formatlara dönüştürmek için birçok yöntem içermektedir ama çizgecik sayımı gibi ileri seviye algoritmalar henüz bu açık kaynaklı yazılım grubuna dahil değildir.

Biyolojik ağ verisinin hacminin büyümesine paralel olarak ağların dinamik doğasının modellenmesi de önemli bir problemdir. Bu amaçla dokulara özgü ağların belirlenmesi için çalışmalar yapılmıştır (Will ve Helms, 2016; Kotlyar vd., 2016). Ama hem fiziksel etkileşimlerin, hem gen düzenleme etkileşimlerinin, hem de metabolik etkileşimlerin entegre edilerek doku ya da hastalık bazında ağların oluşturulduğu bir çalışma bulunmamaktadır.

Çizgecikler bir çizgenin topolojik özellikleri hakkında daha detaylı bilgi verebilen küçük alt-çizgeler olarak tanımlanır. Tek düğümlerle ilgili ortalama derece, derece dağılımı gibi topolojik özelliklere ek olarak, çizgecikler bir düğümün çevresi ile ilgili daha kapsamlı bilgi sağlarlar. İlk olarak 2007 yılında Przulj (2007) tarafından önerilen çizgeciklerin biyolojik ağların analizi ve karşılaştırmasında etkili oldukları gösterilmiştir (Hayes vd., 2013). Fakat olası farklı çizgecik sayılarını düğüm sayısı ile birlikte üssel olarak artması nedeniyle çizgeciklerin hızlı bir şekilde büyük çizgelerde sayılabilmesi zor bir problemdir. 2-5 düğüm büyüklüğünde yönlü veya yönsüz çizgelerin hızlı sayılması için birçok algoritma geliştirilmiştir (Marcus ve Shavitt, 2012; Meira vd., 2014; Hocevar ve Demsar, 2014). Fakat bu yöntemlerin hiçbirisi çok etiketli kenarları gözönüne almamaktadır. Farklı çizgecik sayılarının üssel olarak artması çok etiketli yönlü çizgelerde çok daha hızlı bir şekilde gerçekleşmektedir. Bu projede yönlü ve çok etiketli çizgecikler entegre ağlarda sayılmış ve biyolojik olarak kullanım alanları ve etkinlikleri gösterilmiştir. Bu yönüyle çalışmamız bu alanda bir ilk olmuştur.

3. GEREÇ VE YÖNTEM

Bu bölümde doku ve hastalıklara özel entegre ağların Pathway Commons veritabanı ve NCBI GEO veritabanındaki GSE7307 transkriptom verisi kullanılarak nasıl elde edildiği ve çok etiketli yönlü çizgeciklerin nasıl sayıldığı anlatılmaktadır. Ayrıca çizgecik sayılarının istatistiksel olarak anlamlandırılma yönteminin detayları da verilmektedir. Son olarak iki ağda benzer olan alt ağların etkili görselleştirilebilmesi için geliştirdiğimiz ağ yerleştirme algoritmasının da detayları verilmektedir.

3.1 Entegre Biyolojik Ağların Oluşturulması

Bu proje çalışmasında elde ettiğimiz 141 doku/hastalık entegre ağını oluşturmak için 2 veritabanından faydalandık: 1) Pathway Commons veritabanı ve 2) NCBI GEO veritabanı. Pathway Commons hem uygulamalar hem de çeşitli veriler içeren bir biyolojik ağ portalıdır. Bu projede Pathway Commons'da sunulan SIF formatındaki ağlar kullanılmıştır. SIF formatında elde ettiğimiz bu ağ, Reactome, HPRD ve BioGRID gibi birçok farklı veri kaynağından oluşturulmuştur. Bu entegre ağ 19537 insan proteini arasındaki 883211 etkileşimi içermektedir. Tablo 1'de bu ağda bulunan etkileşim tipleri ve sayıları verilmektedir. Fiziksel etkileşimde olma ve aynı komplekste olma tipindeki etkileşimler yönsüz, diğerleri ise yönlü etkileşimlerdir.

Tablo 1. Pathway Commons'dan elde edilen ağdaki etkileşim tipleri ve sayıları

<i>Etkileşim Tipi</i>	<i>Etkileşim Sayısı</i>
Fiziksel etkileşim	369895
Aynı komplekste bulunma	153603
Fosforilasyonunu kontrol etme	15636
Bir önceki reaksiyonu katalize etme	120948
Gen ifadesini kontrol etme	110013
Taşınmasını kontrol etme	6960
Durum değişikliğini kontrol etme	106156

Fosforilasyon ve taşınmayı kontrol etme etkileşimlerinin tamamı aynı zamanda durum değişikliği etkileşimi olarak da gösterildiği için durum değişikliğinin bir alt kümesi olan bu etkileşim tiplerini kenar etiketi sayısını azaltmak amacıyla gözardı ettik. Sonuç olarak 19537 protein arasında 523498 yönsüz ve 337117 yönlü etkileşim içeren bir ağ elde ettik. Bu ağ hem fiziksel etkileşimler, hem gen düzenleyici etkileşimler, hem de metabolik etkileşimler

içermektedir. İki protein arasında varolan farklı tipteki etkileşimleri tek bir çok etiketli yönlü kenar haline dönüştürerek gösterdik ve bunun sonucunda 760,754 protein ikilisi arasında toplam 1,521,508 yönlü etkileşim olan bir entegre ağ elde etmiş olduk. Bu gösterimde yönsüz etkileşimlere ek olarak yönlü etkileşimler de iki yönlü etkileşim olarak kenar üzerinde gösterilmiştir. Yani A proteini B proteininin durumunu kontrol ediyorsa, A'dan B'ye olan "durumunu kontrol eder" kenarına ek olarak B'den A'ya da "durumu kontrol edilir" etkileşimi eklenmektedir. Bu kenarları ikilik düzende bir kodlama ile göstererek kenar örüntülerinin sayılmasını gerçekleştirdik. Bu yöntemin detayları bir sonraki alt bölümde anlatılmaktadır.

Belirli dokularda ve hastalık durumlarında aktif olan genleri belirlemek için bir büyük ölçekli bir transkriptom profillemeye deney verisi kullandık. Aktif genler listesini kullanarak ilk adımda elde ettiğimiz ana çizgenin aktif olan genleri içeren alt çizgesini kullanarak o hastalığa ya da dokuya ait olan ağı elde etmiş olduk. Aktif genleri belirlemek için GSE7307 mikrodizi veri setini kullandık. Bu veri setinde 90 farklı doku tipinden toplam 677 örnek analiz edilmiştir. Bazı doku tipleri hastalık durumlarını da içerdiğinden farklı doku ve hastalıkları gösteren toplam 141 örnek tipi bulunmaktadır. RMA ile normalize edilmiş gen ifade değerlerini kullanarak aktif genleri belirledik. Aktiflik olarak düşük bir eşik değeri kullanarak parlaklık değeri 10.0 üzerinde olan bütün genleri aktif olarak kabul ettik. Bunun sonucunda 11931'den (retrocervical infiltrate normal örnek tipi) 15853'e (MDA-MB231 kontrol örnek tipi) değişen düğüm sayılarında toplam 141 tane entegre ağ elde ettik.

3.2 Yönlü Çizgeciklerin Sayılması

Bir entegre biyolojik ağdaki protein çiftler arasındaki etiketli etkileşim listesi verildiğinde her bir proteinden diğerine olan yönlü çok etiketli etkileşimi 8 bit bir vektör olarak kodlayarak çizgeyi bir yakınlık listesi (İng. adjacency list) olarak gösterdik. Bu kodlamada A ve B proteinleri arasındaki bir kenarın, ilk iki biti yönsüz etkileşimleri göstermekte, devam eden üç biti A'dan B'ye olan yönlü etkileşimleri ve son üç biti de B'den A'ya olan yönlü etkileşimleri göstermektedir. Tablo 2'de hangi etkileşim tipinin bu kodlamada hangi bite atandığı gösterilmektedir.

Tablo 2. İki protein arasındaki farklı etkileşimlerin kodlanması

<i>Bit</i>	<i>Etkileşim Tipi</i>
1	Fiziksel etkileşim
2	Aynı komplekste bulunma
3	Bir önceki reaksiyonu katalize etme
4	Gen ifadesini kontrol etme
5	Durum değişikliğini kontrol etme
6	Bir sonraki bir reaksiyonu katalize etme
7	Gen ifadesinin kontrol edilmesi
8	Durum değişikliğinin kontrol edilmesi

Örnek olarak A ve B proteinleri arasında aşağıda yer alan etkileşimler verildiğinde:

A B aynı komplekste

A B gen ifadesini kontrol etme

B A bir önceki reaksiyonu katalize etme

yakınlık listesinde A'dan B'ya olan kenar 01010100 olarak kodlanmakta ve B'den A'ya olan kenar da 01100010 olarak kodlanmaktadır. Bu kodlama sisteminde her bir proteinin yakınlık listesi komşuları ile olan bütün etkileşim bilgilerini tutmaktadır. Hash tablosu tabanlı bir sayma yöntemi ile ağdaki bütün çizgecikleri saydık. Bu projede 2 ve 3 düğüm içeren bütün çizgecikleri saydık. Daha büyük çizgecikler için hash tablosu tabanlı sayma yöntemi yetersiz kalmaktadır ve projeyi takip edecek araştırmalar tarafından daha büyük çizgeciklerin hızlı sayılması için yöntemler geliştirilebilir. Bu projede biz 2 ve 3 düğümlü çizgeciklerin farklı tür kenarlar içeren bir entegre ağı karakterize etmede ve farklı doku ve hastalıklar arasındaki farkları ortaya çıkarmada etkili olduğunu gösterdik.

İki düğümlü bir çizgecik basit olarak iki protein arasındaki 8 bit ile kodlanmış kenar vektörü olarak gösterilebilir. Bu vektörü standart bir hash tablosu veri yapısında anahtar olarak kullanıp entegre ağdaki bütün farklı 2 düğümlü çizgecikleri bütün kenarların üzerinden geçerek sayabiliriz. Fakat dikkat edilmesi gereken bir husus A'dan B'ye olan kenar ve B'den A'ya olan kenarın simetrik fakat birbirlerinden farklı kodlanması nedeniyle bu tür kenarların sayısının hash tablosunda iki kere ayrı ayrı anahtarlarda sayıldığı ve bu sayının yanlışlıkla iki kez çıktığı olarak verilmemesidir. Bu izomorfik çizgecikler çizgecik sayılarını çıktığı olarak verirken göze alınmakta ve izomorfizm testi uygulanarak her bir izomorfik çizgecikler grubu

için bir tane çizgecik sayısı belirtilmektedir. Analiz ettiğimiz 141 entegre ağda 56 farklı tip iki düğümlü çizgecik bulunmaktadır.

A, B ve C proteinleri arasındaki üç düğümlü bir çizgeciği A-B, B-C ve C-A kenarlarının 8 bit kodlarının ardışık olarak yazıldığı 24 bitlik bir vektör olarak gösterdik. Sayma sırasında entegre ağın kenar kümesinde bütün ardışık kenarları oluşturduk ve A-B, B-C kenarlarını elde ettik. C ve A arasında bir kenar olmaması durumunda ilgili çizgeciğin son 8 bitini 00000000 olarak belirttik. Aynı iki düğümlü çizgecikleri saydığımız gibi 24 bitlik vektörü hash tablosunun anahtarı olarak kullanarak entegre ağdaki bütün olası üç düğümlü çizgecikleri saydık. Bu sayma yöntemi çok etiketli bir çizgeciğin en spesifik bulunduğu çizgeciği saymak için kullanılmaktadır. Yani çok etiketli bir kenarın alt küme etiketleri daha genel diğer çizgeciklerin sayısına bir katkıda bulunmamaktadırlar. Bu da zaten hedeflediğimiz bir davranıştır. İzomorfik çizgecikleri yine aynı şekilde çizgecik sayılarını çıktı olarak verirken göz önüne alıp her bir izomorfik çizge grubu için bir tane çizgecik sayısını çıktı olarak verdik. Aynı zamanda, üç düğümlü çizgecik türleri çok fazla sayıda olduğu için bir eşik değeri uygulayarak 141 ağdan herhangi birinde en az 10 kez geçmeyen üç düğümlü çizgecikleri gözardı ettik. Analiz ettiğimiz ağlarda 7346 farklı üç düğümlü çizgecik en az bir ağda 10'dan fazla geçmektedir. En büyük doku/hastalık ağında bütün 2 ve 3 düğümlü çizgecikleri sayma işlemi 1.4 GHz Intel Core i5 işlemcili ve 4 GB hafızalı bir MacBook Air dizüstü bilgisayarda 15 dakikada tamamlanabilmektedir.

3.3 Çizgeciklerin İstatistiksel Olarak Anlamlandırılması

Belirli bir çizgecik sayısının z-puanını o çizgecik sayısını, gerçek ağlarla aynı derece dağılımına sahip bir rastgele ağlar kümesindeki çizgecik sayıları ile karşılaştırarak hesapladık. Çok etiketli ve yönlü bir çizgenin derece dağılımı bir düğümden çıkan ya da ona doğru olan kenarların hepsinin gözönüne alan çok boyutlu bir dağılımdır. 141 doku ağında en sık geçen çok etiketli kenar 10000000 kodlu fiziksel etkileşim kenarındır. Bu kenar ortalama olarak 542829,58 kere geçmektedir. En az sıklıkla bulunan çok etiketli kenar ise 00101101 kodlu (bir önceki reaksiyonu kataliz etme, durum değişikliğini kontrol etme, bir sonraki reaksiyonu kataliz etme ve durum değişikliğinin kontrol edilmesi) kenarındır. Bu kenar da ortalama 10,83 kez geçmektedir. 56 farklı çok etiketli kenardan 12 tanesi, 141 ağ içinde ortalama 1000 ya da daha fazla geçmektedir. 28 tanesi ise 100 ya da daha fazla geçmektedir.

141 doku/hastalık ağının her birini kenar karıştırma yöntemi ile rastsallaştırdık. Aynı çoklu etikete sahip rastgele seçilmiş iki kenarı etkileşen proteinler yer değiştirecek şekilde değiştirdik. Örnek olarak, eğer A-B ve C-D kenarları aynı etikete yani aynı 8 bitlik koda sahip kenarlar ise, ağdan A-B ve C-D kenarlarını çıkartıp A-D ve C-B kenarlarını ekledik. Bu işlem sırasında A, B, C ve D proteinlerinin birbirlerinden farklı proteinler olduğuna dikkat ederek rastsallaştırma sırasında öz-kenarlar oluşmasını engellemiş olduk. Bütün kenarları bu şekilde karıştırdıktan sonra bu 141 ağdaki iki ve üç düğümlü çizgeleri saydık. Bir çizgecik için rastgele çizgelerdeki ortalama çizgecik sayısını, m_g , ve standard sapmasını, s_g , hesapladık. Ayrıca çizgecik sayılarının bu rastgele ağlarda normal dağılımda olduğunu kontrol ettik ve daha sonra gerçek bir doku ya da hastalık ağında c_g kez geçen bir g çizgeciğinin z-puanını aşağıdaki denklem ile hesapladık:

$$Z_g = (c_g - m_g) / s_g$$

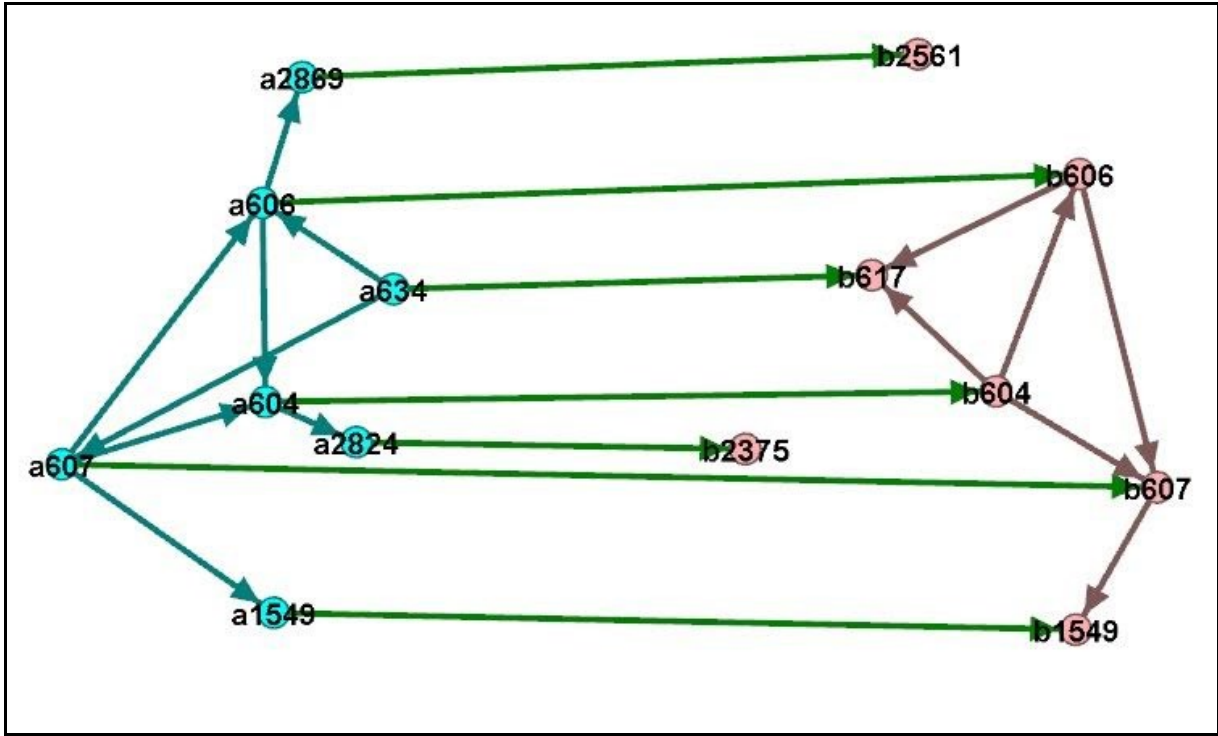
Eğer bir çizgecik rastgele ağların hiçbirinde geçmiyorsa yani m_g ve s_g 0 ise o zaman o çizgeciğın z-puanını tanımsız olarak belirledik.

3.4 Çizgelerde Eşleşen Bölgelerin Etkili Görselleştirilmesi

Doku veya hastalığa özgü iki çizge arasında eşleşen bir alt-çizgenin anlamlı bir şekilde görselleştirilebilmesi varolan çizge yerleştirme (İng. graph layout) yöntemleri ile mümkün olmamaktadır; çünkü varolan yöntemler tek bir çizge için tasarlandıklarından farklı çizgelerin düğümleri karışık olarak yerleştirilmektedir. Bu sorunu çözmek için bizim önerdiğimiz yöntemde farklı çizgelerin düğümleri $x=0$ doğrusunun solunda ve sağında yer alacak şekilde öncelikle yerleştirilirler. Burada çözmemiz gereken bir problem hangi düğümlerin eşleştiğini belirten eşleşme kenarlarının birbiri ile kesişerek algılamayı zorlaştırmamasıdır. Bunun için eşleşen düğümler aynı y koordinatına sahip olacak şekilde soldaki ya da sağdaki çizgede bir sıralama yapılabilir. Bu sıralama için herhangi bir permutasyon kullanıldığında bu sefer çizgelerin kendi içlerindeki kenarlar birbirleri ile çok sayıda kesişebilirler ve bu da çizgelerin eşleşen alt çizgelerinin birbirlerine ne kadar benzediğinin algılanmasını zorlaştırır. Birbirleri arasında kenar olan düğümlerin permutasyonda birbirlerine yakın olacak şekilde bir sıralama bulunması NP-tam bir problem olan minimum lineer ayarlama (İng. Minimum Linear Arrangement) problemi ile aynı problemdir. Bu nedenle bu projede minimum lineer ayarlama problemi için önerilmiş yaklaşık çözümlerden birisi olan Koren ve Harel (2002) yönteminden faydalandık. İlk olarak birinci çizgenin düğümlerini yukarıdan aşağıya, düğümler rastgele sıralanacak şekilde sıraladık. İkinci çizgenin düğümlerini de birinci çizgeyi referans alarak eşleşen düğümler karşı karşıya gelecek şekilde sıraladık. Daha sonra her bir çizge için

sırayla düğüm pozisyonlarının hesaplandığı bir yöntem izledik. İlk önce düğümlerin y koordinatlarını hesaplayıp daha sonra x koordinatlarını belirledik. y koordinatlarını belirlemek için Koren ve Harel'in önerdiği çözümde ayarlamaların maliyetini azaltmak amacıyla kullandıkları medyan tekrarlama (İng. median iteration) yöntemini kullandık. Bu yöntemde her bir düğüm için komşularını ve kendisini gözönüne alıp bunların o anki iterasyondaki y koordinatlarının medyanını hesapladık ve o düğümü bu medyan pozisyonuna taşıdık. Bu işlemi belirli bir sayıda tekrar ederek birbirine bağlı olan düğümlerin yakın y koordinatlarında yer almalarını sağlamış olduk. x koordinatları için derecesi yüksek olan düğümlerin x=0 doğrusunda daha uzakta olmasını sağlamak için onların x koordinatlarını daha solda ya da daha sağda olacak şekilde dereceye bağlı olarak düzenledik.

Şekil 1'de birbirleri ile eşleşen iki alt-çizge için örnek bir çizge yerleşim sonucu verilmektedir.



Şekil 1. Eşleşen alt-çizgelerin önerilen çizge yerleştirme yöntemi ile görselleştirilmesi

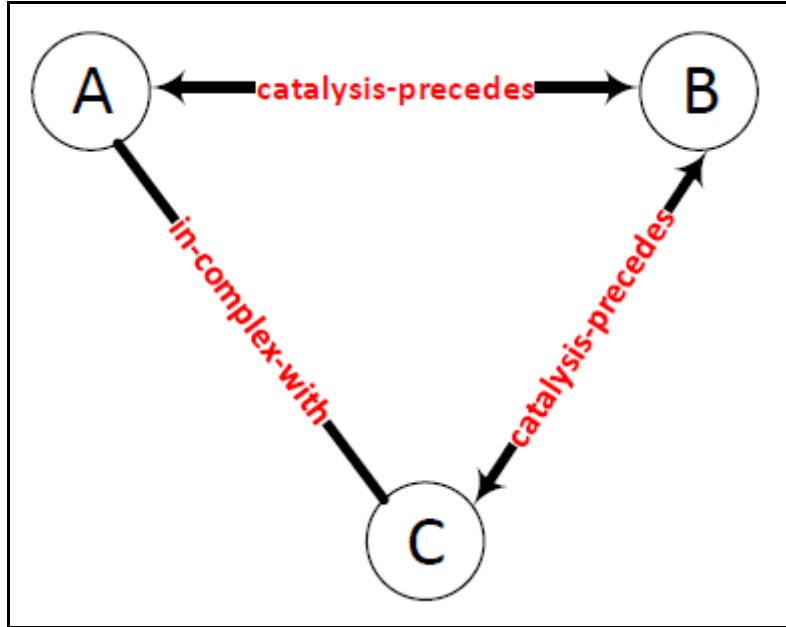
4. BULGULAR

Yönlü ve çok etiketli çizgecik sayılarının olası kullanım alanlarını üç ayrı problem altında örnekledik. İlk olarak büyü hastalık ve doku ağlarında istatistiksel olarak en anlamlı çıkan çizgecikleri detaylı olarak inceledik. Daha sonra farklı doku ya da hastalıkların gruplanması için bu çizgeciklerin nasıl kullanılabileceğini irdeledik. Son olarak da iki durum arasında (örnek: hastalıklı doku, sağlıklı doku) farklı sayılarda olan çizgecikleri inceledik. Aşağıda alt bölümlerde elde ettiğimiz sonuçlar detaylı olarak anlatılmaktadır.

4.1 İstatistiksel Olarak Anlamalı Çizgecikler

İncelediğimiz 141 doku/hastalık ağında en anlamı çıkan çizgecikleri bu bölümde detaylandırdık. Rastgele ağların hiçbirinde olmayan çizgecikler için bir z-puanı hesaplanamadığından bunları gözardı ettik. Bu şekilde 2766 çizgecik vardı. Kalan 4636 çizgecikten 275 tanesi 100.00'in üzerinde bir z-puanı elde etti.

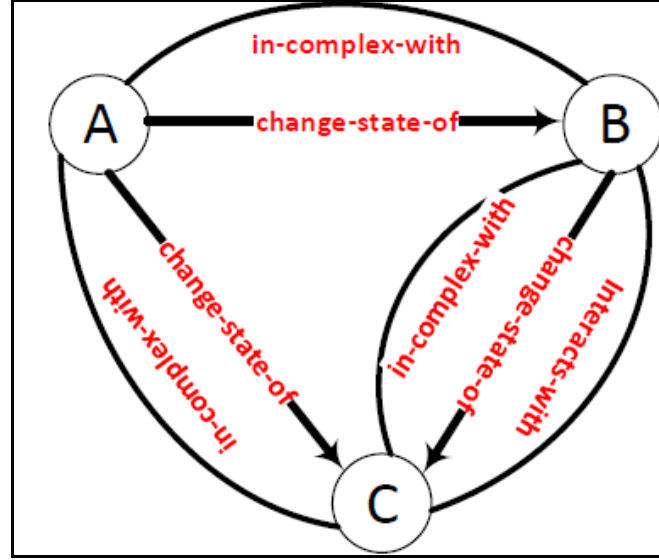
Şekil 2'de 141 doku ağı içinde ortalama 6913,36 z-puanına sahip, istatistiksel olarak en anlamlı çizgecik verilmektedir. 141 rastgele ağda bu çizgecik ortalama 10,35 kere geçmesine rağmen gerçek ağlarda ortalama elli binden fazla geçmektedir. Bu çizgecikte hem metabolik hem de protein kompleks etkileşimleri yer almaktadır. Biyolojik olarak ilginç olabilecek bir gözlem ise metabolik reaksiyonlarda birbirinin hem öncesinde hem de sonrasında yer alan proteinlerin çokluğudur.



Şekil 2. 141 doku/hastalık ağında istatistiksel olarak en anlamlı çizgecik

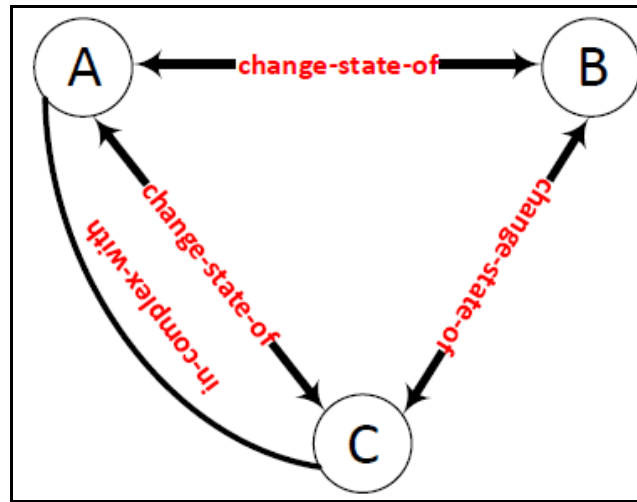
İstatistiksel olarak anlamlı çizgeciklerin çok olması nedeniyle, biyolojik olarak ilginç olan anlamlı çizgecikleri incelemek için z-puanı 141 ağ içinde oldukça değişim gösteren anlamlı çizgeciklere baktık. Şekil 3'te 141 doku/hastalık ağı içerisinde en fazla z-puanı varyansına sahip olan çizgecik görülmektedir. Ortalama z-puanı 4656,97 olan bu çizgeciğin z-puanı standart sapması 1570,49'dur. Bu çizgecik üç protein içeren bir protein kompleksidir ve durum değiştirme ile ilgili olan etkileşimler bir ileri besleme döngüsü (İng. feed-forward loop) motifi oluşturmaktadırlar. Bu çizgecik myometriyum hastalık dokusunda (MDT) 7999 z-puanı

ile en fazla geçmektedir. Aynı çizgecik retroservikal infiltrate normal dokusunda (RINT) 1702,81 z-puanı ile yaklaşık olarak 4 kez daha az geçmektedir.



Şekil 3. En fazla z-puanı varyansına sahip anlamlı çizgecik

Doku ağları içinde yüksek z-puanı varyansına sahip olan bir başka ilginç çizgecik Şekil 4'te verilmektedir. Bu çizgecikte, durum değiştirme etkileşimleri bütün üç protein arasında birbirlerinin durumlarını değiştirme etkileşimi olarak göze çarpmakta, ayrıca bu proteinlerden ikisi birbirleri ile bir kompleks oluşturmuş durumdadır. Bu çizgecik endometriyum yumurtalık hastalık dokusunda (EODT) 7564 kez geçmesine rağmen iskelet kas dokusunda (SMSQNT) 1378 kez geçmektedir. Bu çizgeciğin rastgele ağlarda ortamala bulunma miktarı 2.0'dır.

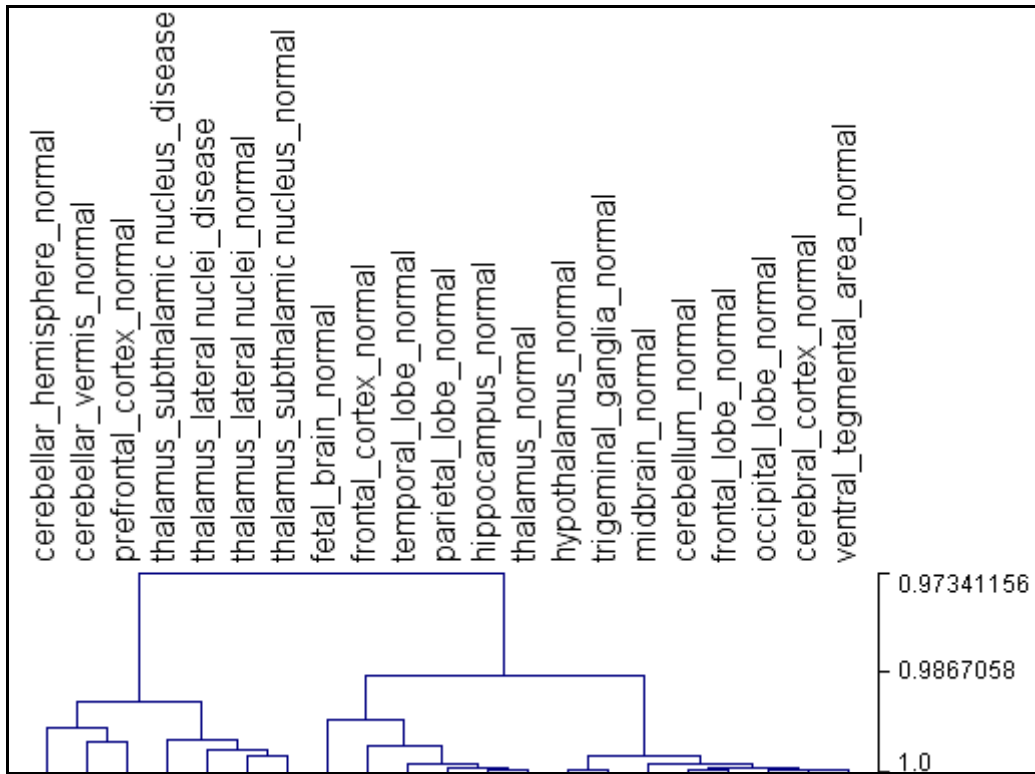


Şekil 4. Yüksek z-puanı varyansına sahip bir başka anlamlı çizgecik

4.2 Doku ve Hastalık Ağlarının Çizgecik İmzaları Kullanılarak Kümelenmesi

Çizgeciklerin z-puanlarını kullanarak (yine bir önceki bölümde anlatıldığı gibi z-puanı hesaplanamayan çizgecikleri gözardı ederek) her bir doku/hastalık ağını çizgeciklerin z-puanlarından oluşan bir vektör olarak gösterdik. Bu şekilde bir vektör gösteriminden sonra birçok farklı kümeleme, sınıflandırma yöntemleri kullanılarak bu ağlar analiz edilebilir. Bu bölümde biz, hiyerarşik kümeleme ve temel bileşen analizi (İng. Principal Component Analysis) (PCA) yöntemlerini uygulamamız sonucunda elde ettiğimiz bulguları sunduk. Bu analizler için TM4 MeV: MultiExperiment Viewer versiyon 4.9 (Saeed vd., 2003) yazılımını kullandık. Bu yazılım aslında mikrodizin analizleri için tasarlanmış bir yazılım olmasına rağmen PCA ve hiyerarşik kümeleme gibi standart yöntemlerin vektör veriler üzerinde uygulanmasını içerdiğinden bizim yaptığımız analizler için uygun bir yazılımdır.

Her bir 141 ağ için elde ettiğimiz çizgecik imzaları üzerinde PCA uyguladığımızda ilk üç temel bileşen 141 ağdaki varyansın %96,126'sını açıklamak için yeterli oldu. Bu sonuç, her ne kadar ağlarda binlerce farklı çizgecik olsa da, farklı çizgecik sayılarının birbirleri ile oldukça ilişkili olduğunu ve ağların aslında çok daha az boyutlu bir vektör ile gösterilebileceğini vurgulamaktadır.



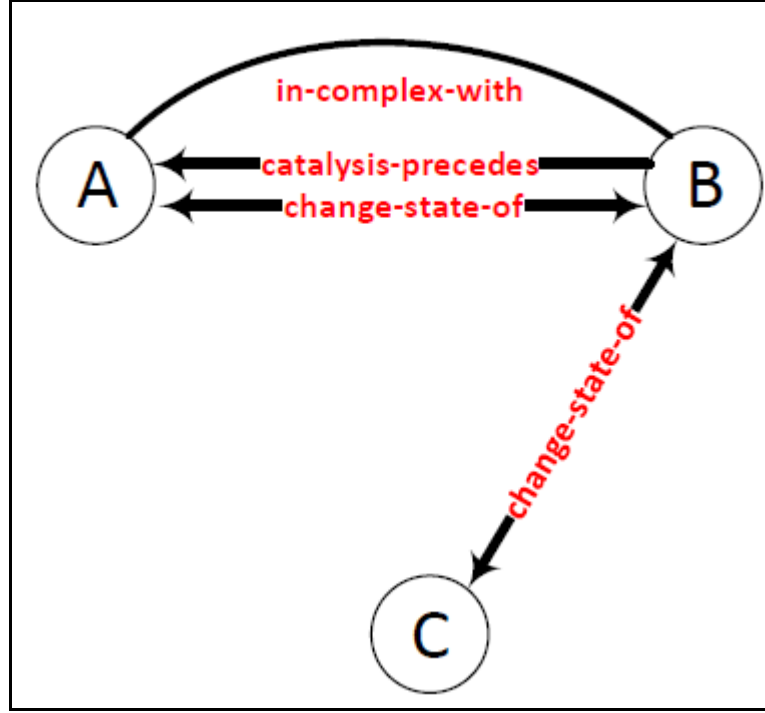
Şekil 5. 21 beyin dokusunun çizgecik imzaları kullanılarak hiyerarşik kümelenmesi

Çizgecik imzaları kullanarak ağların kümelenebileceğini göstermek için 141 ağ içerisinde beyin dokusu ile ilgili olan 21 tane doku ağını çizgecik vektörleri kullanarak kümeledik. Şekil 5'te doku ağlarının hiyerarşik kümelenmesinin sonucu verilmektedir. Şekilden görüleceği üzere benzer fonksiyonlara sahip beyin alt dokular birlikte kümelenmişlerdir. Eğer hiyerarşik sınıflandırma dört küme oluşak şekilde kesilirse, beş talamus dokusundan dördü aynı kümede yer almaktadır. Aynı kümeleme işlemini çizgecik vektörleri yerine GSE7307'deki gen ifade verileri ile yaptığımızda talamus dokuları farklı kümelere düşmektedir. Bu da çizgecik vektörleri ile ağ gösteriminin kümelemede gen ifade verilerine göre daha etkili olduğunu göstermektedir. Ek olarak, ağlardaki aktif gen sayısı gibi basit ağ özellikleri de çizgeciklerin oluşturduğu kümelemeden farklı kümelemeler üretmektedir. Özet olarak, bu sonuçlar farklı tür dokuların kümelenmesi ve karşılaştırılmasında çizgecik vektörleri tabanlı gösterimin etkili bir gösterim olduğunu ortaya koymaktadır.

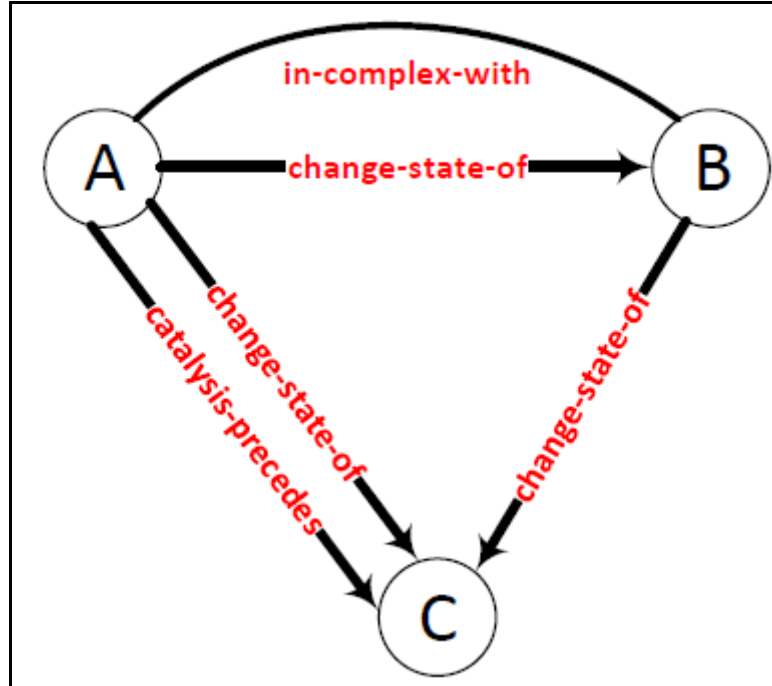
4.3 İki Biyolojik Durum Arasında Farklı Sayıda Olan Çizgeciklerin Belirlenmesi

Hastalık mekanizmalarını anlamada faydalı olabilecek bir uygulama olarak normal ve hastalıklı dokular arasında farklı sayıda geçen çizgecikleri araştırdık. Çizgecik z-puan imzalarını hastalıklı ve normal dokular arasında inceleyerek bu tür çizgecikleri belirledik. Hem hastalık hem de normal örnekler içeren talamus lateral çekirdek, prostat bezi ve deri dokusu örneklerinde bu analizlerimizi gerçekleştirdik. İki doku arasında farklı sayıda geçen çizgeciklerdeki proteinleri de bir fonksiyonel zenginleştirme yazılımı olan DAVID (Huang vd., 2008) ile inceledik ve bu farklara neden olan proteinlerde öne çıkan biyolojik fonksiyonları belirledik.

Şekil 6'da talamus lateral çekirdek dokusu normal ve hastalıklı dokular arasında en fazla z-puanı değişimi gösteren çizgecik sunulmuştur. Bu çizgeciğin hastalıklı dokudaki z-puanı -0.08'dir. Yani rastgele ağlardaki ortalama sayıdan bile daha az sayıda bu çizgecik hastalıklı dokuda geçmektedir. Bu çizgeciğin normal dokudaki z-puanı 43,40'tır. Hastalık dokusunda bu çizgeciği oluşturan sadece 6 tane farklı protein bulunmasına rağmen normal dokuda aktif olan 43 protein bu çizgeciği oluşturmaktadırlar. DAVID ile iki ağ arasında bu farklılığa yol açan proteinlerde öne çıkan biyolojik fonksiyonlara baktığımızda bir çoğunun büyüme faktörü aktivitesi ve hücre bölünmesinin aktivasyonu ile alakalı olduğu gözükmektedir. Yani hastalıklı dokuda bu proteinlerin aktive olmaması bu fonksiyonların gerçekleştirilemediğinin ve dolayısıyla hastalık mekanizmasının bir göstergesi olabilir.



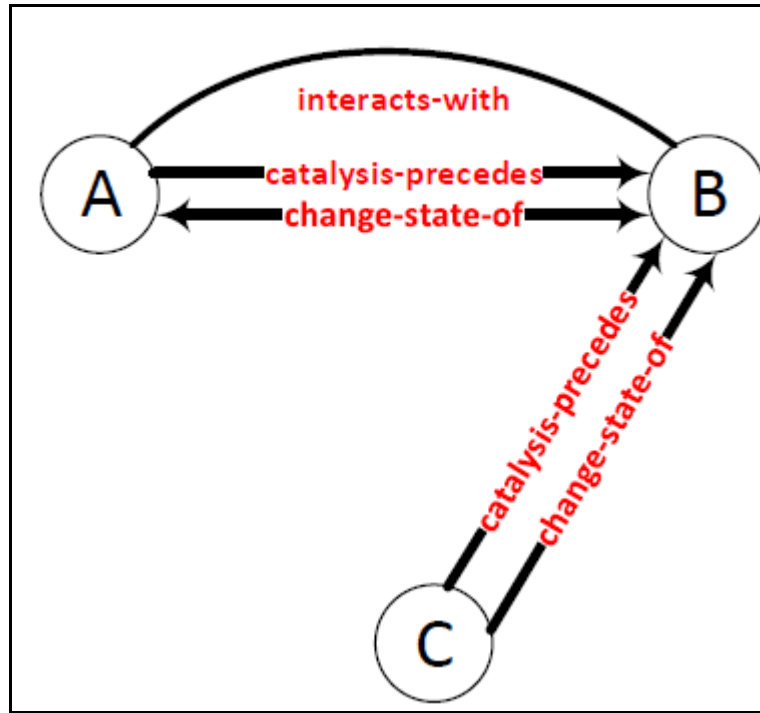
Şekil 6. Talamus lateral çekirdek dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik



Şekil 7. Prostat bezi dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik

Şekil 7'de de prostat bezi dokusunda normal ve sağlıklı dokuda farklı sayıda geçen bir çizgecik gösterilmektedir. Bu çizgeciğin normal dokudaki z-puanı 7,63 iken, hastalıklı dokudaki z-puanı 68,68'dir. Bu çizgecik hastalıklı dokuda normale oranla yaklaşık 9 kat daha fazla geçmektedir. Hastalıklı dokuda bu çizgeciği ortaya çıkartan 74 proteine karşılık normal dokuda bu 74 proteinin yalnızca 28 tanesi aktif olarak gözükmemektedir. Yalnızca hastalıklı dokuda aktif olan 46 proteinin fonksiyonlarını DAVID ile incelediğimizde bu proteinlerin transmembran alıcı protein tirozin kinaz sinyalleme yolağı, büyüme faktörü aktivitesi ve yine hücre bölünmesinin aktivasyonu fonksiyonları ile öne çıktığını görmekteyiz.

Hastalık ve normal karşılaştırmasının son örneği olarak normal ve hastalıklı deri dokularının ağlarını karşılaştırdık. Şekil 8'de yalnızca normal deri dokusu ağında bulunan ve hastalık doku ağında hiç bulunmayan bir çizgecik gösterilmiştir. Normal doku ağında bu çizgeciği oluşturan aktif 69 proteinin fonksiyonların incelediğimizde bunların çoğunun transmembran transport aktivitesi proteinleri olduğunu gördük. Bu fonksiyonun yitirilmesi sonucu ortaya çıkmış olabilecek deri hastalığı mekanizmasını daha detaylı biyolojik incelemeler sonucunda netleştirmek mümkün olabilir.



Şekil 8. Deri dokusu hastalıklı ve normal örnekte farklı sayıda olan bir çizgecik

5. TARTIŞMA VE SONUÇ

Bu proje kapsamında yapılan çalışmada, istatistiksel olarak anlamlı olan çizgeciklerin dokuları sınıflandırma ve hasta ve normal dokular arasında sistematik farkları belirleme gibi farklı biyolojik problemlerdeki potansiyel faydaları gösterilmiştir. Normal ve hastalıklı dokularda farklı sayılarda geçen çizgeciklerin biyo-belirteç olarak kullanılabilmesi için bu farkların arkasında yatan biyolojik nedenler laboratuvar deneyleri ile daha detaylı incelenmeli ve doğrulanmalıdır. Bu projede önerilen ve gerçekleştirilen çizgecik sayma yöntemi geliştirilebilecek en hızlı yöntem değildir ve bu nedenle üçten fazla düğüm içeren çizgeciklerin sayımını etkili olarak gerçekleştirememektedir. Bu projeden sonra gelecekte yapılabilecek çalışma olarak çok etiketli yönlü çizgeciklerin daha hızlı sayılması için yöntemler geliştirilebilir. Bu projede genom-ölçekli entegre ağların karakterizasyonu için ilk kez yönlü ve çok etiketli çizgeciklerin sayılması önerilmiştir. Bunun gerçekleştirilebilmesi için bit-kodları tanımlanmış ve hash tablosu tabanlı bir yöntem önerilmiştir. Önerilen yöntemi oldukça geniş ölçekli 141 örnek içeren doku/hastalık ağlarında test ettiğimizde hem istatistiksel olarak anlamlı çizgecikler bulduk hem de bu çizgecik vektörleri tabanlı gösterimin dokuları gen ifade verilerine kıyasla daha etkili bir şekilde kümelediğini gösterdik. Bu çizgeciklerin, moleküler kompleksler, sinyalleme yolları ve metabolik yollar gibi daha büyük biyolojik süreçler içindeki görevlerinin ortaya konabilmesi yine proje devamında yapılacak işler arasında listelenebilir.

6. KAYNAKLAR

- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N. vd. 2011. "Pathway Commons, a web resource for biological pathway data", *Nucleic Acids Research*. 39(suppl 1):D685-D690.
- Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. 2007. "Integration of biological networks and gene expression data using Cytoscape", *Nat. Protoc.*, 2, 2366-2382.
- Demir, E., Cary, M., Paley, S., vd. 2010. "BioPAX - A community standard for pathway data sharing", *Nature biotechnology*. 28(9):935-942.
- Demir, E., Babur, O., Rodchenkov, I., Aksoy, B.A., Fukuda, K.I., Gross, B., vd. 2013. "Using Biological Pathway Data with Paxtools", *PLoS Comput Biol*. 9(9):1-5.
- Hayes, W., Sun, K., Przulj, N. 2013. "Graphlet-based measures are suitable for biological network comparison", *Bioinformatics*. 29(4):483-491.
- Hocevar, T., Demsar, J. 2014. "A combinatorial approach to graphlet counting", *Bioinformatics*. 30(4):559-565.
- Huang, D.V., Sherman, B.T., Lempicki, R.A. 2008. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources", *Nature Protocols*. 4:44-57.
- Kamburov, A., Wierling C., Lehrach, H., Herwig, R. 2009. "ConsensusPathDB - a database for integrating human functional interaction networks", *Nucleic Acids Resarch*, 37:623-628.
- Kamburov, A., Stelzl U., Lehrach, H., Herwig, R. 2013. "ConsensusPathDB interaction database: 2013 update", *Nucleic Acids Resarch*, 41:D793-D800.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., vd. 2009. "Human Protein Reference Database - 2009 update", *Nucleic Acids Research*. 37(suppl 1):D767-D772.
- Koren, Y., Harel, D. 2002. "A Multi-scale Algorithm for the Linear Arrangement Problem", *Graph-Theoretic Concepts in Computer Science: 28th International Workshop, WG 2002 Cesky Krumlov, Çek Cumhuriyeti, 13-15 Haziran*, 296-309.
- Kotlyar, M., Pastrello, C., Sheahan, N., Jurisica, I. 2016. "Integrated interactions database: tissue-specific view of the human and model organism interactomes", *Nucleic Acids Research*. 44(D1):D536-D541.
- Maraziotis, I.A., Dimitrakopoulou, K., Bezerianos, A. 2007. "Growing functional modules from a seed protein via integration of protein interaction and gene expression data", *BMC Bioinformatics*, 8, 408.
- Marcus, D., Shavitt, Y. 2012 "RAGE - A rapid graphlet enumerator for large networks", *Computer Networks*. 56(2):810-819.
- Meira, L.A.A., Maximo, V.R., Fazenda, A.L., Da Conceicao, A.F. 2014. "Acc-Motif: Accelerated Network Motif Detection", *IEEE/ACM Trans Comput Biol Bioinformatics*. 11(5):853-862.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. 2002. "Network Motifs: Simple building Blocks of Complex Networks", *Science*, 298:824-827.

Przulj, N. 2007. "Biological network comparison using graphlet degree distribution", *Bioinformatics*. 23(2):e177-e183.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., vd. 2003 "TM4: a free, open-source system for microarray data management and analysis", *Biotechniques*. 34:374-378.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M. 2006. "BioGRID: a general repository for interaction datasets", *Nucleic Acids Research*. 34(suppl 1):D535-D539.

Will, T., Helms, V. 2016. "PPIXpress: construction of condition-specific protein interaction networks based on transcript expression", *Bioinformatics*. 32(4):571-578.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., Margalit, H. 2004. "Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction", *Proc. Natl Acad. Sci.*, 101:5934-5939.

7. EKLER LİSTESİ

Ek-1 Dergi Makalesi. Arzu Burçak Sönmez ve Tolga Can, "Comparison of tissue/disease specific integrated networks using directed graphlet signatures," BMC Bioinformatics, kabul edildi, basım aşamasında.

Ek-2 Konferans Bildirisi. Arzu Burçak Sönmez ve Tolga Can, "Comparison of tissue/disease specific integrated networks using directed graphlet signatures," BCB '16 Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 533-534, Seattle, WA, USA — Ekim 02 - 05, 2016.

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje Yürütücüsü:	Doç. Dr. TOLGA CAN
Proje No:	114E111
Proje Başlığı:	Farklı Nitelikteki Biyolojik Ağların Entegrasyonu Ve Yerel Topolojik Özellik Vektörleri Tabanlı Karşılaştırılması
Proje Türü:	1001 - Araştırma
Proje Süresi:	24
Araştırmacılar:	
Danışmanlar:	YEŞİM AYDIN SON
Projenin Yürütüldüğü Kuruluş ve Adresi:	ORTA DOĞU TEKNİK Ü. MÜHENDİSLİK F. BİLGİSAYAR MÜHENDİSLİĞİ B.
Projenin Başlangıç ve Bitiş Tarihleri:	01/12/2014 - 01/12/2016
Onaylanan Bütçe:	102124.0
Harcanan Bütçe:	43697.87
Öz:	<p>Bu projede, entegre genom-ölçekli ağların yönlü çizgecik imzaları kullanılarak analiz edilmesi için yöntemler geliştirilmiştir. Ek olarak, hizalanmış ağların etkili görselleştirilebilmesi için bir ağ yerleştirme yöntemi önerilmiştir.</p> <p>Büyük hacimli verilerin heterojenliği nedeniyle entegre genom-ölçekli ağların analizi zor bir araştırma problemidir. Biyolojik ağların karakterizasyonu için çizgecik sayıları gibi birçok topolojik ölçüt bulunmaktadır. Bu projede, etiketli ve çok yönlü çizgelerdeki küçük alt-çizge örüntülerini saymak için yöntemler önerilmiştir. Proje kapsamında analiz ettiğimiz ağları Pathway Commons veritabanından elde ettiğimiz ve H. sapiens genlerini kapsayan fiziksel etkileşimleri, gen düzenleyici etkileşimleri ve metabolik etkileşimleri içeren ağları entegre ederek elde ettik. Bu entegre ağı daha sonra büyük ölçekli bir transkriptom profillemeye çalışmasını kullanarak doku ve hastalık bazında filtreledik ve farklı doku ve hastalıklar için özel entegre ağlar elde etmiş olduk. Yönsüz protein-protein etkileşim ağları için önerilmiş olan çizgecik sayma fikrini geliştirerek çok yönlü ve etiketli ağlara uyguladık ve her bir ağ için çizgecik sayı vektörlerini oluşturduk. Bu çizgecik sayılarının istatistiksel anlamlandırmasını rastgele ağlardan elde ettiğimiz çizgecik sayıları ile karşılaştırarak gerçekleştirdik. Çizgecik sayılarının etkinliğini farklı doku ve hastalıklardaki çizgecik sayılarını karşılaştırarak ve farklı durumlardaki ağları kümeleyerek gösterdik. Elde ettiğimiz sonuçlar entegre biyolojik ağlarda çok sayıda istatistiksel anlamlı çizgeciğin bulunduğunu ve ağların kümelenmesi ve sistem seviyesinde doku ve hastalık ağlarının karşılaştırılması için çizgecik sayılarının etkili bir gösterim olduğunu ortaya koymaktadır. Ek olarak, önerilen çizge yerleştirme algoritmasının karşılaştırılan ağlardaki benzerlik ve farklılıkların görselleştirilmesi için etkin bir yöntem olduğu gösterilmiştir.</p>
Anahtar Kelimeler:	entegre biyolojik ağlar, ağ karşılaştırma, yönlü çizgecikler, hizalanmış ağların görselleştirilmesi
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır