



Biyolojik Ağların Gaussian Grafiksel Modellerle Tahmininde Kopulaların Uygulanması ve Parametre Tahmini

Program Kodu: 1001

Proje No: 114E636

Proje Yürütücüsü:
Doç. Dr. Vilda PURUTÇUOĞLU

Araştırmacı(lar):
Dr. Tamay ŞEKER

Danışman(lar):
Prof. Dr. Ernst WIT
Doç. Dr. Alexandre De Leon

Bursiyer(ler):
Hajar FARNOUDKIA
Deniz SEÇİLMİŞ

MAYIS 2017
ANKARA

Önsöz

Çalışma biyolojik ağların modellemesinde karşılaşılan iki temel problem olan yüksek boyutlu olma ve seyreklik oranının fazla olması sorunlarının çözümüne yönelik olarak, kopula Gaussian grafiksel modelinin (Copula Gaussian graphical model – CGGM) kullanılmasına ve bu modelin parametrelerinin tahminin Bayesci metotlardan biri olan ters atlamalı Markov zinciri Monte Carlo (reversible jump Markov chain Monte Carlo – RJMCMC) yöntemiyle yapılmasına yöneliktir. Çalışmada hem modelleme aşamasında hem de önerilen modelin farklı dağılımlı verilerin analizinde değerlendirilmesi sırasında kopulalar uygulanmış ve farklı doğruluk kriterleriyle gerçek ve simüle edilmiş veriler üzerinden önerilen model ve alternatif yöntemler kıyaslanmıştır. Çalışma TÜBİTAK tarafından desteklenmektedir.

İçindekiler

Şekil Listesi.....	III
Özet.....	IV
Abstract.....	V
Giriş.....	1
Literatür Özeti.....	3
Gereç ve Yöntem.....	10
Bulgular.....	12
Tartışma ve Sonuç.....	16
Kaynaklar.....	18

Şekil Listesi

Şekil 1.....	1
Şekil 2.....	4

Özet

Proje gerçekçi büyüklükteki karmaşık biyolojik ağ yapılarının sistem davranışlarını durağan haldeyken modellenmesi ve model parametrelerinin Bayesci yaklaşımlarla tahmin edilmesini kapsamaktadır. Modellemede kopula Gauassin grafiksel modeli (CGGM) kullanılmış ve parametre tahmini ise ters atlamalı Markov zinciri Monte Carlo yöntemiyle yapılmıştır. Burada önerilen modelin literatürde sıklıkla kullanılan Gaussian grafiksel modeline (GGM) göre avantajı, sistemin durumunu açıklayan yüksek boyutlu birleşik çok değişkenli normal dağılımının Gaussian kopula sayesinde marjinallerine ve kovaryans matrisine ayrılabilmesidir. Biyolojik sistemlerin yüksek boyutlu olması ve sıklıkla bu tür verilerde gözlem sayısının, parametre sayısına göre az olması sebebiyle de, model parametre tahmininde Bayesci yaklaşımlar tercih edilmiştir. Sonuçlardan önerilen modelin ve Bayesci algoritmanın mevcut yöntem ve algoritmalarından daha yüksek doğruluğa sahip olduğu bulunmuştur. Diğer yandan hesaplama süresinin uzun olduğu gözlemlenmiştir. Bu sorunu çözmek amacıyla da proje önerisinin dışında farklı alternatifler araştırılmış ve tahminin Gibbs örneklem metoduyla da yapılabileceği görülmüştür. Ayrıca normallik varsayımına dayanan GGM'in aslında normal dışındaki durumlarda doğruluk oranının oldukça düştüğü farklı kopulaya dayalı analizlerle görülmüştür. Bu sebeple de yine proje önerisi dışında alternatif yeni modellerin geliştirilmesi için BAP projesi yazılmış ve bu proje ODTÜ (Orta Doğu Teknik Üniversitesi) BAP ofisi tarafından kabul edilmiştir. Bu sebeple projenin ilerleyen aşamalarında bulgular sadece GGM ile değil, önerilen yeni modellerle de kopula verileri yardımıyla kıyaslanmıştır.

Anahtar Kelimeler: Kopula Gaussian grafiksel modeli, ters atlamalı Markov zinciri Monte Carlo algoritması, kopula, matematiksel modelleme, sistem biyoloji.

Abstract

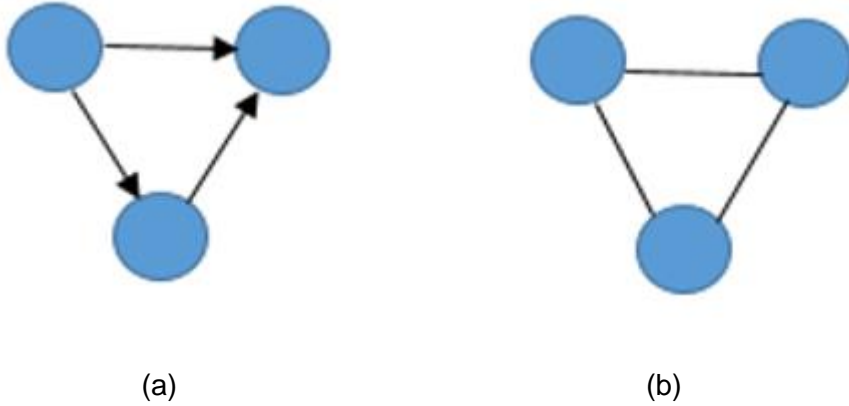
The project covers the modellings of the system's behaviour of the realistic complex biological networks under the steady-state condition and the inference of the model parameters via the Bayesian approaches. In modelling, the copula Gaussian graphical model (CGGM) has been used and the estimation of parameters has been conducted by the reversible jump Markov chain Monte Carlo method. The advantage of the suggested model over the Gaussian graphical model (GGM), which is frequently used in the literature, is to separate the high dimensional joint multivariate normal distribution into the marginals and the covariance matrix by means of the Gaussian copula. As the biological networks have high dimension and the number of observations are generally less than the number of parameters in this type of data, the Bayesian approaches have been preferred in inference of the model parameters. From the results, it has been found that the proposal model and the Bayesian algorithm have higher accuracy than the existing methods and algorithms. On the other hand, it has been observed that the computational time is demanding. Hereby, in order to solve the underlying challenge, different alternatives have been studied beyond the content of the project and it has been seen that the Gibbs sampling can be applicable for this purpose. Moreover, from the findings of the distinct copula analyses, it has been detected that the rate of accuracy of GGM, which is based on the normality assumption, decreases significantly under non-normal cases. Hence, for this purpose, a new BAP (scientific research project) project has been written and this project has been accepted by the METU (Middle East Technical University) BAP Office. Accordingly, in the continue of the project, the findings have not been only compared with GGM, but also with new proposed models by means of the copula measurements.

Keywords: Copula Gaussian graphical model, reversible jump Markov chain Monte Carlo algorithm, copula, mathematical modelling, systems biology.

1. GİRİŞ

Herhangi bir sistemin elemanlarının etkileşimini gösteren yapılara *ağ* veya *sistem* adı verilmektedir. Kullanımları sosyal bilimlerden biyolojik bilimlerine ve mühendisliklere kadar çok yaygın olduğu için, ağ yapıları kullanıldıkları alanın ismiyle beraber anılır. Fakat tüm ağ yapılarının ortak özelliği, hepsinin de düğüm (nodes) ve kenarlar (edges) adı verilen parçalardan oluştuğudur. Düğümler, sistemin elemanlarını, kenarlar da düğümler arası etkileşimi temsil etmektedir. Dolayısıyla biyolojik ağlarda düğümler *genleri* veya *türleri*, kenarlar ise genler arası *etkileşim* veya *aktivasyonu* ifade etmektedir. Grafik teorisinde, düğümler çemberle, kenarlar ise çizgilerle gösterilir (Whittaker, 1990).

Biyolojik ağlar, çok farklı organizma yapılarını içerdiği için, sıklıkla, düğümleri ifade eden sistem elemanlarının türlerine göre adlandırılır. Buna göre biyolojik ağları moleküler, genomik ve proteomik ağlar olarak üçe ayırabiliriz. Metabolik ağlar *moleküler ağlara*; gen düzenleyici ağlar *genomik profile*, genlerden sentezlenen proteinlerin fonksiyonu ve birbirleri ile etkileşimi ise *proteomik ağlara* örnek olarak verilebilir. Bir de çok genel anlamda ve tüm farklı bilimlerdeki ağ yapılarını da kapsayacak biçimde ağlar, *yönlü* (directed) ve *yönsüz* (undirected) ağlar olmak üzere sınıflandırılabilir (Junker ve Schreiber, 2008). Yönlü ağlar, düğümler arası ilişkinin yönünü, kenarlarda oklarla belirten ağ çeşididir. Yönsüz ağlar ise sadece ilişkilerin varlığını ifade etmektedir. Şekil 1'de her iki ağ yapısı türü gösterilmektedir.



Şekil 1. (a) Yönlü ve (b) yönsüz ağ yapısı gösterimi.

İşte belirtilen tüm bu farklı ağ elemanlarını, durumlarını ve birbirleriyle olan etkileşimlerini ifade edebilmek için biyolojik ağların matematiksel modellemesinden yararlanılmaktadır. Bu sebeple matematiksel modelleme, bir ağın veya sistemin anlaşılabilmesi ve analiz

edilebilmesi için kullanılan en temel yoldur. Bir denklem olarak ifade edilebilen sistemler için, böylece, hem sistemin büyük ve karmaşık yapısı açıklanabilir, biyolojik hipotezleri test edilebilir ve sistem elemanlarının davranışlarıyla ilgili tahmin yapılabilir duruma gelmektedir. Diğer yandan sistem aktivasyonunu düğüm ve kenarlar cinsinden görsel olarak da ifade edebilmemize imkan sağladığı için, sistem elemanlarını ve onların birbirleriyle ilişkilerini grafiksel gösterimlerle özetlemek de mümkün olabilmektedir (Wilkinson, 2006).

Tüm bunlara dayanarak, biyolojik ağların matematiksel modellemesinde iki temel yaklaşım kullanılır. Bunlar *deterministik* ve *stokastik* yaklaşımlardır. Deterministik yaklaşımlar, diğer bir deyişle modeller, sistemleri, temel fizik kuralları kullanarak tanımlar. Bu modellerde sistem, durağan haldeki koşullarına göre değerlendirildiği için sistemin verilen bir t zamanındaki durumu ve bir sonraki ölçüm noktasındaki davranışı, tamamen belirlenebilir. *Boolean* ve *adi diferansiyel denklemler* bu modelleme grubuna girmektedir. *Stokastik yaklaşımlar* ise sistemin rassal davranışlarını açıklamak için kullanılır ve sistemin bu sayede dinamik yapısını açıklar. Bu sebeple de stokastik modellerde sistem elemanlarının davranışları, olasılık dağılımlarıyla açıklanır. Dolayısıyla biyolojik sistemlerin modellemesi, temelde *boolean*, *adi diferansiyel denklemler* ve *stokastik modeller* olmak üzere üç ana başlıkta toplanabilir (Bower ve Bolouri, 2001). Bunların içinde boolean yaklaşımı, biyolojik ağların modellenmesinde kullanılan en basit modeldir ve sistem durumlarını (states) ifade eden genlerin aktivasyonunu, sadece aktif veya aktif değil olarak temsil eder. Bunu yaparken de ikili mantığı (0-1) kullanır. Ayrıca sistemin mevcut ve sonraki durumunu anlatırken basit mantık kurallarına dayalı bir boolean fonksiyonu tanımlar ve bu fonksiyon deterministiktir. Bu basit yapısı sebebiyle de biyolojik sistemlerin karmaşık yapılarını açıklamak için yetersizdir ve sıklıkla yeni araştırılan sistemler için ön bilgi edinmek amacıyla kullanılır (Bolouri, 2008).

Adi diferansiyel denklemler modeli ise gen durumlarının tanımlanmasını genlerin yoğunluklarını kullanarak yapan deterministik bir modellemedir. Bu yoğunluklar sürekli ve zamana bağlıdır ve sistemin zamana bağlı değişimi reaksiyon hız sabitine göre gerçekleşir. Adi diferansiyel denklemler de model parametreleri olan bu reaksiyon hızlarını hesaplamak için kullanılır. Ancak kimyasal reaksiyon sayısı arttıkça, diğer bir deyişle sistem büyüdükçe, diferansiyel denklemler doğrusal olmayan denklemlere dönüşürler. Bu da denklem sisteminin çözümünü zorlaştırmaktadır. Dolayısıyla büyük ve karmaşık sistemlerde model parametrelerinin tahmini sorunlu olmaktadır. Fakat modelleme için gerekli olan gözlemlerin elde edilmesi ve sistem hakkında verdiği detay sebebiyle, en çok tercih edilen modelleme tekniğidir (Jong, 2002).

Üçüncü dal olan stokastik modeller ise, kullandığı bilginin detayı sebebiyle, boolean ve diferansiyel denklem modellerine göre daha kapsamlı modellerdir. Bu yaklaşımda, sistem açıklanırken, genin aktif olup olmama durumu, genlerin molekül sayılarındaki değişimle ifade edilmektedir. Dolayısıyla modelde, kesitli niceliklerle belirtilirken, hangi değişimin ne zaman olacağı kimyasal ana denkleminde (chemical master equations) (Van Kampen, 1981; Gillespie, 1977) dayanarak olasılıksal olarak anlatılmaktadır. Bu modelleme türü her ne kadar diğer alternatif iki ana model yapısına göre sistem hakkında daha detaylı bilgi verebilse de, modellemede gerekli olan girdi bilgilerini elde etmek, özellikle gerçekçi büyüklükteki sistemler için oldukça zordur. Hesaplamaların yapılabilmesi için gerekli bu bilgi kısıtı sebebiyle de kullanımı yaygın değildir. Bu nedenle de pratik anlamda Langevin, difüzyon modeli gibi yaklaşık sonuç veren stokastik modeller daha çok kullanılır (Vyshemirsky vd., 2004; Endy ve Brent, 2001; Wilkinson, 2006; Purutçuoğlu ve Wit, 2008; Purutçuoğlu, 2012).

Belirtilen tüm bu sebeplerden dolayı, bu projede karmaşık sistemlerin modellemesinde, sistemin durağan halini açıklamaya yönelik modelleme türü ele alınmıştır. Projede önerilen metod, diferansiyel denklemlerin olasılıksal dengi görülebilecek *Gaussian grafiksel modeli* (Gaussian graphical model – GGM) ve bu modelin yine büyük ve karmaşık sistemlerde görülen tahmin sorununun çözümüne yönelik olarak düşünülebilecek kopula Gaussian grafiksel modeli (copula GGM – CGGM) olarak belirlenmiştir.

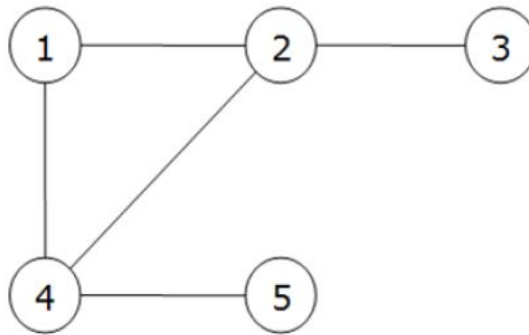
2. Literatür Özeti

Sistem elemanları arasındaki ilişkiyi ifade eden ağlar, kabul ettikleri varsayımlara göre farklı biçimlerde matematiksel olarak ifade edilebilir. Bu modeller en temel anlamda Boolean, diferansiyel denklemlere ve stokastik model yapılarıdır (Bower ve Bolouri, 2001). Alternatifler içerisinde Boolean, ağın aktivasyonuna ait sadece çalışır/çalışmaz halinde bilgi verdiği için, özellikle literatürde yeterli bilginin bulunmadığı ağları anlamak için kullanılır. Stokastik modeller ise, Boolean'ın aksine ağ ile ilgili detaylı bilgi elde etmede tercih edilir ve sistemin rassal yapısını açıklamada kullanılır. Bu tür aktivasyonlar protein transkripsiyonu veya translasyonu işlemlerinde görülür ve ortamda sınırlı sayıda molekül varsa gözlemlenebilir. Dolayısıyla, modellemede, sistemdeki her bir elemanın (gen veya protein) ölçüm anındaki molekül sayılarına ihtiyaç duyar. Fakat bu modelleme, hem veri toplama konusunda maliyetli ekipmanlara ihtiyaç duyması sebebiyle, hem de sistem hakkında detaylı bilgiyi modellemede kullanması sebebiyle, sadece çok iyi bilinen ve çalışılan ağlar için tercih edilmektedir (Wilkinson, 2006; Golightly ve Wilkinson, 2011; Purutçuoğlu ve Wit, 2008; Purutçuoğlu, 2012; Boys vd., 2008). Buna karşın adi diferansiyel denklem

modellemeleri sistemin durağan haldeki aktivasyonu hakkında bilgi verdiği için sistemin genel yapısını anlamda tercih edilir. Ayrıca mevcut verilerin bir çoğu bu tür modellemeye uygundur (Bower ve Bolouri, 2001; Jong, 2002). Ancak adi diferansiyel denklem modellemesi optimizasyon yöntemlerine dayalı tahmin yapar ve deterministik olarak ağ yapısını açıklar (Defterli vd., 2014). Bu sebeple de ağ yapısı büyüdükçe ve seyrekleştikçe (sparsity), tahmin zorlaşmakta ve model parametre tahminleri, global yerine yerel (local) optimal sonuçları vermeye başlamaktadır. Diğer yandan bu tür modellemeye alternatif olarak önerilen Gaussian grafiksel modeli (GGM) ise, ağın durağan halini olasılıksal olarak açıklar (Whittaker, 1990; Defterli vd., 2014). Bu nedenle de hem dağılım bilgisini hem de rassallığı katmaya imkan sağladığı için, model parametre tahminlerinde daha fazla esneklik tanımaktadır. Bu sebeple, hazırlanan bu çalışmada da, GGM temel modelleme çeşidi olarak tercih edilmiştir.

GGM, yönsüz grafiklere dayanan yaygın bir olasılıksal modelleme yöntemidir. Bu model, sistem elemanları olan genlerin Y , $Y \sim N(\mu, \Sigma)$ parametreleri ile çok değişkenli Gaussian (Normal) dağılımından geldiği varsayımına dayanmaktadır. Burada sistemdeki toplam gen sayısını n ile gösterirsek, ortalama $\mu = (\mu_1, \dots, \mu_n)$ ve varyans-kovaryans matrisi Σ , $(n \times n)$ boyutla ifade edilebilir. GGM'de temel kolaylık, genler arasındaki ilişkinin koşullu bağımsızlık bilgisi ile anlatılabilmesidir (Whittaker, 1990; Defterli vd., 2014). Buna göre örneğin Şekil 2'de verilen beş genden oluşan bir sistem için, gen 2 değerinin bilinmesi şartıyla, gen1 ve gen 3 veya gen 2 ve gen 3 birbirinden koşullu olarak bağımsızdır. Bu bağımsızlık durumu GGM'de modelin de temel matematiksel ifadesi olan ve Denklem (1)'de de gösterilen lasso regresyonundaki regresyon katsayısı olan β ile orantılıdır. Nitekim β , kovaryans matrisinin tersi olan kesinlik matrisi (precision matrix) $\theta = \Sigma^{-1} = \theta_{ij}$ ile de gösterilebilir. Bu sebeple GGM'de model parametresi olarak θ alınır ve bu sayede Denklem (1)'deki β ile ilişki kurulur.

$$Y^{(P)} = Y^{(-p)}\beta + \varepsilon \quad (1)$$



Şekil 2. 5 genden oluşan yönsüz basit bir sistem yapısı.

Denklem (1)'de $Y^{(p)}$, p . genin durumunu, $Y^{(-p)}$, ise p ($p=1, \dots, n$) geni dışındaki diğer sistem genlerini anlatmaktadır. Diğer bir deyişle $Y=(Y^{(p)}, Y^{(-p)})'$ dir ve $Y^{(-p)} = (Y^{(1)}, \dots, Y^{(p-1)})$ ile son gen dışındaki tüm diğer genlerden oluşan vektörü ifade etmektedir. Bu nedenle $Y^{(p)}$ geninin diğer tüm genler cinsinden şartlı (conditional) dağılımını, aşağıda sunulan dağılımla göstermek mümkündür. Son olarak ε , normal dağılımdan gelen rassal hatayı göstermektedir. Ayrıca bu ifadede $(.)'$ terimin transpozunu, μ ve Σ ise aşağıda verilen şekliyle yazılabilir.

$$Y^{(p)} | Y^{(-p)} = y \sim N(\mu_p + (y - \mu_{-p})' \Sigma_{-p,-p}^{-1} \sigma_{-p,p}, \sigma_{p,p} - \sigma_{-p,p}' \Sigma_{-p,-p}^{-1} \sigma_{-p,p}). \quad (2)$$

$$\mu = \begin{bmatrix} \mu_{-p} \\ \mu_p \end{bmatrix} \text{ ve } \Sigma = \begin{bmatrix} \Sigma_{-p,-p} & \sigma_{-p,p} \\ \sigma_{-p,p}' & \sigma_{p,p} \end{bmatrix}. \quad (3)$$

Denklem (3)'de μ_{-p} , son gen dışındaki tüm diğer genlerin ortalama değerini ve μ_p , son genin ortalamasını göstermektedir. Diğer yandan $\Sigma_{-p,-p}$, son gen dışında kalan ve $((p-1) \times (p-1))$ boyutlu kalan genlerin varyans-kovaryans matrisini, $\sigma_{-p,p}$, $((p-1) \times 1)$ boyutlu $Y^{(-p)}$ ile $Y^{(p)}$ arasındaki kovaryans vektörünü ve son olarak $\sigma_{p,p}$, son gen $Y^{(p)}$ 'nin varyansını belirtmektedir. Literatürde seyreklik (sparsity) durumu altında, Denklem (1)'deki β , dolayısıyla $\theta = \Sigma^{-1}$ 'i tahmin etmek için farklı metotlar uygulanmaktadır. Bunlardan başlıcaları *küçültme* (shrinkage) yöntemi, komşuluk algoritmalarında dayalı *lasso-tabanlı* yöntemler (Meinshausen ve Bühlmann, 2006) ve *grafiksel lasso* algoritmasıdır (Friedman vd., 2008). Bunlardan *küçültme yöntemi*, sistemin gerçek yapısı bilinip de genler arasındaki etkileşimin kuvveti hakkında tahminde bulunulması gereken hallerde kullanılan bir metottur. Bu sebeple karmaşık sistem modellemelerindeki kullanımı kısıtlıdır. Diğer yandan *lasso-tabanlı yöntemlerde* sistemin doğru yapısının bilinmediği kabul edilir ve bu özelliği ile daha çok seyrek yapıları ağların tahmin edilmesinde kullanılmaktadır. Dolayısıyla amaç hem sistemin yapısını tahmin etmek hem de genler arasında tahmin edilen etkileşimin kuvvetini belirlemektir. Üçüncü metot olan grafiksel lasso, veya kısaca *glasso*, metodunda ise yüksek seyrekliğe sahip kesinlik matrisi tahmin etmek için kesinlik matris elemanlarına L_1 -cezalandırması (L_1 -penalty) metodu (Tibshirani, 1996) uygulanmaktadır (Friedman vd., 2008). Bu metodun uygulaması, diğer ikisine göre daha yaygın olsa da, her üç metodun genel problemi olan yüksek boyutlu sistemlerde hesaplama süresinin uzunluğu ve doğruluk değerlerinde düşme sorunlarını çözememektedir. Bu sorunların çözümüne yönelik benzeri cezalandırma yöntemleri ve en az kareler yöntemi ile kesinlik matrisinin bulunulması çalışılmışsa da, yine büyük sistemler için etkili bir yol bulunamamıştır. **Bunlara örnek olarak “adaptif lasso” (Zou, 2006), “kaynaşık lasso” (fused lasso) (Tibshirani vd., 2005) ve “elastik net” (Zou ve Hastie, 2005) metotları gösterilebilir. Bu yöntemlere alternatif**

olarak son yıllarda Liu vd. (2012) tarafından nonparanormal SKEPTIC, Hsieh vd. (2014) tarafından QUIC ve Wang vd. (2016) tarafından FastGGM adı verilen yeni yöntemler de önerilmiştir. Bu metotlardan ilki grafiksel Dantzig seçicisi ve CLIME gibi parametrik olmayan, diğer yandan kovaryans matrisinin, pozitif yarı tanımlı şartını aramayan algoritmalara dayalı hesaplamalara dayalıdır. Metot, GGM'in normallik varsayımına esneklik sağlamasına rağmen, yöntem tamamen optimizasyona dayalıdır ve eksik gözlem veya kısıtlı gözlem olması durumunda modelin doğruluk performansı düşmektedir. Bu sebeple de mikrodizin verileri gibi çok gözlemin olduğu veri kümeleri için önerilmiştir. QUIC (Quadratic Approximation for Sparse Inverse Covariance Estimation) adı verilen diğer yöntem ise tamamen parametrik olmayan bir yöntemdir ve kesinlik matrisinin tahminini sayısal çözümlerle (numeric solution) bulmaktadır. Yöntem diğer yöntemlere göre oldukça hızlı olsa da hiç bir parametrik bilgi kullanmaması sebebiyle, verinin dağılım bilgisini gözardı etmektedir. Son olarak önerilen FastGGM yaklaşımı ise temelde en az kareler yönetime dayanmakla birlikte, kodlamaların C++ programlama diliyle yazılması ve hesaplamaların hızlanmasına yönelik bazı ara steplerin ilave edilmesi sebebiyle, hesaplama süresinde belirgin bir iyileştirme sunmaktadır. Ancak bu ve diğer bahsedilen metotların tamamında tahmin metodu cezalandırma yöntemlerine dayalı olduğu için, en iyi modele ait cezalandırma sabitinin (penalizing constant) seçimi kullanılan model seçim kriterlerine göre değişmektedir. Bu konuda en çok uygulanan seçim kriterleri k-çarpık doğrulama (k-cross validation) (Friedman vd., 2008), STARS (düzenlilik seçiminde kararlılık yaklaşımı – stability section to regularization selection) (Liu vd., 2010) ve RIC (rotation information criterion - rotasyon bilgi kriteri) (Zhou ve Shen, 2012) olarak sayılabilir. Bunların yanı sıra bu katsayının optimal değerini bulmak için hatalı keşif oranını (false discovery rate – FDR) kullanan metotlar da mevcuttur (Liu, 2013). Dolayısıyla model performansı GGM'deki kullanılan parametrik olmayan veya en çok olabilirliğe dayalı metotların beraberinde kullanılan model seçim kriterlerinden de etkilenmektedir.

Diğer yandan, Dobra ve Lenkoski (2011) tarafından uygulanan *kopula GGM modeli* (CGGM), bu sorunlara sahip olmayan bir sosyal veri analizi için yaptıkları çalışmalarında, model parametresi olan kesinlik matrisini, Bayesci algoritmalarından olan *ters dönüşümlü Markov zinciri Monte Carlo algoritmasıyla* (reversible jump Markov chain Monte Carlo – RJMCMC) tahmin etmişlerdir. CGGM modeli kısaca, GGM modelinin Gaussian kopula (Genest ve Favre, 2007; Trivedi ve Zimmer, 2005) ile birleştirilmiş hali olarak görülebilir. Model çok değişkenli birleşik normal dağılımının marjinaler ve Gaussian kopula terimi ile

parçalanabilmesine imkan sağlamakta, böylece kesinlik matrisi θ 'yı ayrı tahmin etmeyi mümkün kılabilmektedir. Dobra ve Lenkoski (2011) çalışmalarında bu tahmini yaparken kullandıkları RJMCMC metodu aslında bir çeşit Metropolis-Hasting (MH) algoritmasıdır. Bu algoritma literatürde sıklıkla kullanılan temel bayesci algoritmalarından biridir (Gamerman ve Hedibert, 2006; O'Hagan ve Forster, 2004). Ayrıca bu algoritma, asimetrik çekirdek dağılımı ile yapılan atlama kuralları (jumping rules) sayesinde daha hızlı sonsal (posterior) dağılıma yakınsamaktadır (Gelman vd., 2004). Dolayısıyla RJMCMC'de parametre tahmininde sadece parametreler değil, aynı zamanda bunların boyutları da tahmin edilerek bir Markov zinciri yaratılır (Gelman vd., 2004). Diğer bir deyişle algoritma, sistemin parametrelerini güncellerken, bunların da üzerinde parametrelerin geldiği modeli de güncellemektedir. Yani metod, her iterasyonda k tane alternatif sistem modeli arasından en uygun olanı ve ona ait olan parametreleri, θ_k , belirlemektedir. Bu durumda sistemin güncellenmesi için kabul olasılığı, aşağıda verilen ifade ile hesaplanır. Bu değer, standart tek düze dağılımdan ($U(0,1)$) gelen rassal değerden küçük olursa sistem, önerilen güncellemeyi kabul etmez ve eski halinde kalır. Aksi takdirde önerilen yeni sistem yapısı ve ilgili parametre değerleri, sistem için kabul edilir.

$$\alpha = \frac{p(Y | \theta_{k^*}, M_{k^*}) p(\theta_{k^*} | M_{k^*}) \pi_{k^*} J(u^* | k^*, k, \theta_{k^*}) \left| \frac{\nabla(\theta_{k^*}, u^*)}{\nabla(\theta_k, u)} \right|}{p(Y | \theta_k, M_k) p(\theta_k | M_k) \pi_k J(u | k, k^*, \theta_k)} \quad (4)$$

Denklem (4)'deki ifadede π_{k^*} ve π_k terimleri sırasıyla önerilen ve kabul edilmiş olan k . alternatif model için öncül dağılımları göstermektedir. Ölçeksizlik (scale-free) özelliğine sahip olan biyolojik sistemler için (Barabási ve Oltvai, 2004) bu dağılımların power-law, kesikli (truncated) power-law, genelleştirilmiş pareto dağılımı gibi alternatiflerden biri olduğu önerilse de gerçek verilerde, bunlardan hiçbirinin başarılı sonuç vermediği gözlemlenmiştir (Khanin ve Wit, 2006). Diğer yandan $p(\theta_{k^*} | M_{k^*})$ ve $p(\theta_k | M_k)$ dağılımları, yine sırasıyla, önerilen ve mevcut olan model için model parametrelerinin dağılımını ve benzer şekilde $p(Y | \theta_{k^*}, M_{k^*})$ ve $p(Y | \theta_k, M_k)$ dağılımları önerilen ve mevcut olan k modeli için olabilirlik fonksiyonlarını ifade etmektedir. Son olarak $J(u^* | \cdot) / J(u | \cdot)$ oranı, k modeli için önerilen ve mevcut durumun çekirdek dağılımlarının oranı ve $\left| \frac{\nabla(\theta_{k^*}, u^*)}{\nabla(\theta_k, u_k)} \right|$ mutlak değer oranı ise önerilen ve mevcut modelin kabul edilmesi halinde model ve ilişkili olduğu parametre vektörünün eğimini temsil etmektedir. Bu durumda geri dönüşümlü atlama MCMC algoritmasının grafiksel modellemedeki uygulaması aslında, genler arasındaki etkileşimin bilinmediği durumlarda, her biri birer parametre olarak görülebilecek kesinlik matrisinin girdilerinin hangilerinin sıfır,

hangilerinin sıfırdan farklı değer aldığıının belirlenmesi olarak düşünülebilir. Dolayısıyla bu algoritma, sistemdeki toplam parametre sayısının da belli olmaması halinde kullanılan bir yöntemdir ve literatürdeki biyolojik ağ uygulamaları, Lotka Volterra (Wilkinson, 2006) gibi küçük sistemler için yapılmış (Boys vd., 2008), büyük biyolojik sistemlerin tahmin edilmesinden belirtilen şekilde dönüşümlü olarak kullanılmamıştır. **Diğer yandan RJMCMC algoritmasına alternatif bazı algoritmalar da literatürde önerilmiştir. Bunlar “doğum ve ölüm” (birth-and-death) (Mohammadi ve Wit, 2015), “Cralin ve Chib algoritması”(Carlin ve Chibs 1995) ve “Gibbs algoritmasıdır” (Walker, 2009). Ayrıca GGM modelini, bu çalışmada da önerildiği gibi kopula terimiyle birleştirerek CGGM’i kullanıp, model parametre tahmininde ise parametrik olmayan korelasyon katsayısını, kesinlik matrisine alternatif olarak kullanan yeni bir yöntem de önerilmiştir (He vd., 2017). Bu yöntem tahminlerdeki hata keşif oranını (false discovery rate) asimptotik olarak kontrol etse de ve normal olan ve olmayan veri kümeleri için doğruluk oranını yüksek bulsa da, temelde parametrik olmayan yöntemlere dayalı olduğu için, aslında bu çalışmada önerilen kopula GGM ve tamamen parametrik olan tahmin metotlarına göre ayrı gruptadır.**

İşte bu projede CGGM modeli tamamen parametrik yöntemler kullanılarak, ilk defa biyolojik ağ modellemelerinde kullanılıp, orijinal RJMCC kodları yazılarak, model parametre tahmini yapılmaktadır. Böylece sistemlerin boyutuna bakılmaksızın, verilerin dağılım bilgileri göz önünde tutularak, küçük veya yüksek boyutta Gaussian kopula yardımıyla birleşik dağılım fonksiyonu parçalanabilmekte ve parçalanmış yapı da Bayesci algoritma ile tahmin edilebilmektedir. Diğer yandan RJMCMC’ye alternatif önerilen metotlar içerisinde **“doğum ve ölüm algoritması” biyolojik sistemlere adapte edildiği için, proje araştırmalarında RJMCMC sonuçlarını karşılaştırmada kullanılmıştır. Yine alternatiflerden biri olan Gibbs algoritması ise proje önerisinde yer almamakla beraber, proje ekibince, proje çalışmaları sırasında ilk defa biyolojik sistemlere uygulanmış, RJMCMC süresine göre bazı gerçek sistem uygulamalarında hesaplama süresini kısalttığı bulunmuştur (Purutçuoğlu ve Farnoudkia, 2017).**

Projede, bunların yanı sıra, kopula terimi, GGM modelinin kapsamlı bir performans değerlendirmesinde de kullanılmıştır. **Kopula ifadesi en genel anlamda, çok boyutlu birleşik dağılım fonksiyonlarının, marjinaler ve bunlardan ayrı bir kopula terimi adı verilen bağlantı fonksiyonu yardımıyla yazılabilmesini sağlayan matematiksel bir yoldur (Genest ve Favre, 2007; Nelson, 2006). Örneğin $F(X)$ ve $G(Y)$, (X,Y) ortak**

kümülatif dağılıma sahip olan sürekli rassal değişkenlerin, marjinal dağılımları olsun. Bu ortak dağılım, kopula fonksiyonu C ile gösterilmek üzere ($C: [0,1]^2 \rightarrow [0, 1]$), aşağıda verilen biçimde parçalanabilir.

$$H(X, Y) = C[F(X), G(Y); \delta]. \quad (5)$$

Yukarıdaki denklemde δ , $[-1, 1]$ aralığında tanımlı olan korelasyonu göstermektedir ve H biliniyorsa F , G ve C , tek olarak tanımlanabilmektedir. Literatürde C fonksiyonu ile farklı kopula fonksiyonları tanımlanmıştır. Bu fonksiyon Gaussian, student-t, Gumbel, FGM (Farlie-Gumbel-Morgenstern ailesi), Clayton and Frank kopulaları olabilir. Aralarındaki farklılık tanımlı oldukları veri aralığı ve aykırı gözlem içerip içermemelerinden kaynaklanmaktadır (Trivedi ve Zimmer, 2005; Genest ve Favre, 2007; Nelsen, 2006). Diğer yandan Archimedian ailesine dahil olan Gumbel ve Clayton kopularının yüksek boyutlar için dağılım fonksiyonları açık olarak yazılamamaktadır (Wawrzyniak and Kurowicka, 2006). Student-t kopulası ise Gaussian kopula ile neredeyse aynı sonuçları vermekle beraber, hesaplama açısından daha zordur. Bu projede, farklı kopula türleri, biyolojik sistem yapılarının özellikleri düşünülerek araştırılmış, tercihte ise hem yüksek boyuta uygun olan hem de hesaplama kolaylığı sunan tek seçenek olarak Gaussian kopula bulunmuştur. Gaussian kopula, normal olmayan veri üretmede ve böylece bu verilerle GGM için performans değerlendirmesi yapmada kullanılmıştır. Performans değerlendirmelerinde ayrıca gerçek biyolojik sistem yapıları da incelenmiştir. Bu amaca uygun olarak kanser hastalıklarındaki önemi ve ökaryotlardaki büyüme ve bağışıklık sistemlerini kontrol etmeleri sebebiyle MAPK-ERK (Kolch, 2005) ve JAK-STAT (Maiwald vd., 2010; Xu vd., 2013) yolları kullanılmıştır. Bu yolların çalışmalardaki verileri, gerçekçi girdiler kullanılarak (Purutçuoğlu ve Wit, 2012; Purutçuoğlu vd. 2017) Gillespie algoritmasıyla (Gillespie, 1977) simüle edilmiş, yaratılan verilerden örnek seçilerek modelleme uygulamasında kullanılmak üzere veri kümeleri oluşturulmuştur. Bunların yanı sıra literatür taraması yapılarak “hücre sinyal” verisi adı verilen (Sachs vd., 2005) küçük boyutlu bir veri seti ve Stranger vd. (2007) tarafından toplanan insan gen ekspresyon verileri bulunmuştur. Bu iki veri kümesinin ayrıca Mohammadi ve Wit (2015) çalışmasında da karşılaştırma amaçlı kullanıldığı gözlemlendiğinden, bu projedeki tüm araştırmalarda ve modellemelerde, gerçek veriye sahip biyolojik sistemler için “temel karşılaştırma veri kümeleri” olarak kullanılmıştır. Tüm simüle ve gerçek veri analizlerinde ise model karşılaştırmalarında önerilen farklı “model seçim kriterleri” bulunmaktadır (Purutçuoğlu vd., 2017). Bunlardan en bilinenleri F-ölçeği (F-measure), özgüllük (specificity), kesinlik (precision), doğruluk (accuracy), Matthew korelasyon katsayısı

(Matthew correlation coefficient - MCC) ve yanlış pozitif oranı (false positive rate -FPR) olarak sıralanabilir. Projedeki analizlerde alternatifler arasından F-ölçeği (F-measure), Matthew korelasyon katsayısı (MCC) değerleri birden fazla hata türünü eş zamanlı değerlendirebildiği için, özgüllük ise GGM modellemesinde en doğru sonucu veren kriter olarak bulunduğu için (Purutçuoğlu vd., 2017) temel kriterler olarak tercih edilmiştir. Fakat bazı proje çalışmalarına diğer kriterler de zaman zaman ilave edilmiştir.

3. Gereç ve Yöntem

Bu çalışmada kopula ifadesi hem CGGM modelinde yüksek boyutlu birleşik dağılım fonksiyonu parçalamakta, hem de bundan bağımsız olarak, GGM modelinin aslında hangi tür verilerde daha iyi sonuç verdiğini gözlemlemede, özellikle, farklı marjinallerden yaratılmış ve farklı kopulalarla birleştirilmiş veriler üzerinde, tahminlerin doğruluğunu değerlendirmede kullanılmıştır. CGGM içinde kullanımında, model, sadece Gaussian kopulasını kullanmakta ve RJMCC algoritmasıyla, marjinaler ve kopula matrisine ayrılmış yüksek boyutlu ve çok değişkenli birleşik normal dağılımın kesinlik matrisi olan θ , Bayesci yollarla tahmin edilebilmektedir. Bu sayede yüksek boyutlarda GGM kullanımını kısıtlayan model parametre tahmin metotları sorununa, adı geçen yöntem, yüksek boyutu parçalara ayırıp, ayrı ayrı tahmin ederek ve en sonunda tekrar birleştirerek çözüm getirebilmektedir. Çünkü sistemde θ , varsa eksik gözlemler ve saklı (latent) gözlemler, ayrı olarak ve tekrarlayan bir algoritma ile tahmin edilebilmektedir. Böylece aslında GGM modelinin tamamen olasılıksal ve Bayesci tahmin yolu mümkün olabilmektedir. Çünkü GGM'in en sık kullanılan tahmin metodu olan glasso'da, parametre tahmini, cezalandırmalı olabilirlik (penalized likelihood) metotlarıyla yapıldığı için, boyut büyüdükçe tüm parametre değerleri için sayısal çözümün bulunması sorun olmaya başlamaktadır. İşte GGM'in kopula yardımıyla parçalanmış hali olarak düşünülebilecek CGGM ile bunun yapılması mümkün olmuştur.

Diğer yandan GGM'in dayandığı normallik varsayımının da sonuçlardaki etkisini gözlemek amacıyla Gumbell-Hougaard, Galambos, Hüsler-Reiss gibi farklı parametre aralığı kullanan kopulalar incelenmiş, ancak hem verinin boyutu hem de gözlemlerin bulunduğu aralık değerlendirildiğinde biyolojik veriler için ancak Gaussian ve student-t kopulalarının olabileceği gözlemlenmiştir. Bunlardan student-t kopulası, hesaplamalardaki sorunlu yapısı sebebiyle tercih edilmezken, diğer alternatifler de zaten sadece 6 veya 7 boyutta açık dağılım ifadelerine sahip olduğu için bu çalışmada kullanılmamıştır. Gaussian kopula ise hem her boyutta açık kopula matris ifadesini yazabilmeyi mümkün kılması

sebebiyle hem de biyolojik veriler için uygun gözlem aralığını kapsamamasından dolayı bu çalışmada tercih edilmiş ve tüm simülasyonlarda farklı marjinal dağılımları birleştirmek için yararlanılmıştır. Projede normal dağılıma sahip olmayan simüle verilerinin yaratılmasında, marjinaler student-t, log-normal, üssel ve bunların karma dağılımları olarak alınmış, 50, 100 ve 500 gibi farklı boyutlu sistemler düşünülmüş ve tüm bu gözlem değerleri Gaussian kopula ile bağlanarak, istenilen boyutta çok değişkenli birleşik normal dağılıma sahip veri setleri oluşturulmuştur. Tüm simülasyon analizlerinde Monte Carlo ortalama sonuçları hesaplanmış ve iterasyon sayısı 1000 olarak alınmıştır. Bu veri setlerinde Gaussian grafiksel modeline glasso tahmin metoduyla uygulanıp, F-ölçüsü, özgüllük gibi bir çok doğruluk ölçüsü bahsedilen Monte Carlo iterasyonları sonunda hesaplanmıştır. Bu değerlendirmede student-t ve log-normal marjinal dağılım seçim nedeni, bu dağılımların normal dağılımla aynı ailede olması (uzun kuyruklu simetrik dağılım aileri – long tailed symmetric family of distribution) ve normal olmayan yapıları sebebiyle de, GGM'in dağılım değişimindeki performansını gözlemlemeye fırsat verebilmesidir. Bir diğer alternatif olarak üssel dağılım seçim sebebi ise, sistem verilerinin aslında sinyal ölçümleri olması sebebiyle ve piksellerin sayısal değeri olarak ifade edebileceğimiz bu sinyallerin, orijinal ölçekte sıklıkla üssel dağılım ile ifade edilmesidir. Benzeri uygulama mikrodizin çalışmalarında doğru sinyallerin modellenmesinde de kullanılmaktadır.

Son olarak karma dağılımların yaratılmasındaki temel sebep ise gözlemlerin iki dağılımdan oluşması halinde GGM'in bunu modellemedeki başarısını gözlemlemek olmuştur. Burada karma dağılımlar olarak yarı normal ve yarı üssel dağılım karması, bir de yarı normal ve yarı student-t dağılım karması düşünülmüştür. Tüm bu verilerin yaratılmasında ise biyolojik sistemlerin seyrek ve dolayısıyla ölçeksiz (scale-free) yapıları göz önünde tutulmuş ve simüle edilen sistem verilerinin bu şartı sağlaması ön koşul olarak alınmıştır.

Diğer yandan hem CGGM hem de GGM'in gerçek sistem uygulamalarında iki farklı yol izlenmiştir. Bunlardan ilki sistemin gerçek olup, verilerinin uzun süreli Gillespie algoritması (Gillespie, 1977) çalıştırılarak elde edildiği sistemlerdir. Bu sistemler içinde MAPK-ERK (Kolch vd., 2005) ve JAK-STAT (Maiwald vd., 2010; Xu vd., 2013) sistemleri yer almaktadır. Bu sistemlerden ilki ökaryotik hücrelerde hücre büyümesinin kontrol eden ana yollardan (pathway) biridir. Diğerisi ise hücrenin bağışıklık sistemini kontrol eden yoldur. Projede bu sistemlerin simüle verileri yaratılırken, sistemlerin stokastik oran sabit değerlerinin (stochastic rate constant) hesaplandığı ya da bilindiği çalışmalar taranmış (Purutçuoğlu vd., 2013; Ayyıldız vd., 2014) ve bu değerler Gillespie algoritmasında girdi değerleri olarak verilmiştir. Daha sonra sistemdeki tüm proteinlerin durağan hale geldiği zamana kadar sistemler

stokastik olarak yaratılmış, en sonda ise durağan hale gelince, buralardan her bir protein için gözlem değerleri çekilmiştir. Hazırlanan bu veri seti de model performans analizlerinde kullanılmıştır.

Uygulanan bir diğer gerçek sistem verisi ise gerçek gözleme dayalı ancak, simüle veri sistemlerine göre daha küçük boyuta sahip veriler olmuştur. Projede bu şekilde iki veri seti kullanılmıştır. Bunlardan ilki, hücre sinyal verisi (cell signaling data) olarak adlandırılmakta ve toplam 11 genden oluşmaktadır. Diğerisi ise bir insan gen ekspresyon verisidir ve toplam 100 genden oluşmaktadır. Bu verilere ait diğer biyolojik bilgiler, bir sonraki bölümde sunulan makale/tez/bildiri çalışmalarında yer almaktadır.

4. Bulgular

Projede elde edilen bulgular ve metine dökülmüş halleri aşağıdaki üç başlık altında toplanmıştır:

1. *CGGM modellemesi ve model parametre tahmini*: Proje önerisinde sunulan CGGM modeli küçük, orta büyüklükte ve karmaşık sistemlerde uygulanmış ve model parametre tahminleri RJMCMC algoritmasıyla elde edilmiştir. Daha sonra RJMCMC metodundan ve doğum-ve-ölüm Markov zinciri Monte Carlo (birth-and-death MCMC - BDMCMC) metodundan elde edilen sonuçlarıyla iki ayrı F-ölçüsü hesaplanıp, sonuçlar kıyaslanmıştır. Sonuçlardan RJMCMC algoritmasının BDMCMC algoritmasına göre belirgin derecede sonuçları iyileştirdiği gözlemlenmiştir. Bu çalışma aşağıda sıralanan makale çalışmalarında (a1) ile gösterilmektedir ve çalışma şu an uluslararası SCI indeksi bir dergide değerlendirilme aşamasındadır. Daha sonra RJMCMC algoritmasının hızını yükseltmeye yönelik araştırmalar yapılmış ve bu konuda ayırma-karıştırma (split-merge) yaklaşımının, Carlin ve Chib algoritması ve Gibbs örneklem metodlarının olduğu bulunmuştur. Bu alternatifler içinden Gibbs örneklem metodu seçilerek RJMCMC'de olduğu gibi orijinal R programlama kodları yazılıp, aynı kıyaslama sistemlerine uygulanarak, F-ölçüleri ve hesaplama zamanları kıyaslanmıştır. Sonuçlardan Gibbs örneklem metodunun RJMCMC ile aynı doğruluğa sahip olduğu ve çalışılan sistemlerde hesaplama zamanını azalttığı gözlemlenmiştir. Buradan elde edilen tüm bulgular ayrıca (b1)'de toplanmış ve Acta Physica Polonica Series A adlı (SCI indeksli) dergide makale olarak kabul edilmiştir. Son olarak burada CGGM ve RJMCMC yöntemleri, MAPK-ERK sistemine uygulanıp (c1) sonuçların doğruluğu biyolojik olarak da yorumlanmıştır. Diğer yandan özellikle büyük sistemlerde genler arasındaki yüksek korelasyonun RJMCMC'de kesinlik θ matrisinin

parçalanmasına (decomposition) izin vermemesi sebebiyle, sistemde yüksek bağımlılık olması halinde uygulanabilecek bir çözüm yolu önerilmiş, bu çözümde θ 'nin tekil olması halinde Ayyıldız vd., 2010'daki çalışmalarına benzer olarak, bağımlı ve bağımsız genler iki etapda güncellenerek algoritma çalıştırılmıştır. Bu konudaki çalışma ise (c1)'de sunulan çalışmada tam metin olarak kabul edilmiş ve yayımlanmıştır. **Listelenen bu bulguların elde edilmesinde proje yürütücüsü ve doktora bursiyeri Hajar Farnoudkia beraber çalışmıştır. Ayrıca RJMCMC metodunun kodlama uygulamasında proje danışmanlarından Prof. Dr. Alexander De Leon ile, BDMCMC uygulamaları sırasında ise yine proje danışmanlarından olan Prof. Dr Ernst Wit ile e-posta üzerinden görüşülmüştür. Diğer yandan (d1) kodlu yayında kullanılan MAPK-ERK sistemi yapılan hesaplamaların biyolojik değerlendirilmesinde proje biyoloji alan araştırmacısı Dr. Tamay Şeker'den görüş alınmıştır.**

(a1) Farnoudkia, H., Purutçuoğlu, V. 2017. "Bayesian inference of the sparse Gaussian graphical model in the construction of biological network", (hakem değerlendirmesinde), 1-15.

(b1) Purutçuoğlu, V. , Farnoudkia, H. 2017. "Gibbs sampling in inference of the copula Gaussian graphical model adapted to biological networks", Acta Physica Polonica Series A, 1-12 (Kabul edildi).

(c1) Purutçuoğlu, V., Farnoudkia, H. 2016. "Bayesian inference of deterministic MAPK-ERK pathway via reversible jumps Monte Carlo method", Proceeding of the 8th International Conference on Inverse Problems: Modeling and Simulation (IPMS),

(d1) Purutçuoğlu, V., Farnoudkia, H. 2017. "Bayesian inference of biological networks whose components are linearly dependent", Proceeding of Istanbul International Conference on Progress in Applied Science, 1-7.

2. *GGM model performans değerlendirilmesi*: Projede GGM ifadesinin karmaşık ve farklı dağılımlardan gelen gözlemler altında doğruluk değerlerinin hesaplanması için Monte Carlo çalışmaları yapılmış ve bunların 1000 kere tekrarından oluşan ortalama değerleri hesaplanmıştır. Sonuçlardan GGM'in normallik altında en doğru tahminleri vermesine rağmen, normal dışındaki dağılımlarda (üssel dağılım hariç) doğruluk değerinin çok düştüğü, aynı durumun karma dağılım altında da geçerli olduğunu göstermiştir. Yapılan bu kapsamlı simülasyon analiz sonuçları ve gerçek veri uygulamaları aşağıda sıralanan çalışmalardan (a2)'de detaylı olarak sunulmuş ve bu bulgular Acta Physica Polonica Series A (SCI indeksli) dergisinde makale olarak yayınlanması için kabul edilmiştir. **Bu bulguların elde edilmesinde proje**

yürütücüsü ve proje bursiyeri olmamakla beraber, yüksek lisansını proje yürütücüsü ile beraber yapmış olan Damla Dokuzoğlu ile beraber çalışılmış, teorik konularda proje danışmanlarında Prof. Dr. Alexander De Leon ile e-posta ve Skype yoluyla görüşülmüştür. Biyoloji alan uzmanı araştırmacı Dr. Tamay Şeker'e ise GGM modelinin gerçek veri uygulamaları sırasında sonuçların anlamlılığı konusunda danışılmıştır.

(a2) Dokuzoğlu, D., Purutçuoğlu, H. 2017. "Comprehensive analysis of Gaussian graphical model under different biological networks", Acta Physica Polonica Series A, 1-10 (Kabul edildi).

3. *GGM' e alternatif modellerin yaratılması ve kopula uygulamaları:* Proje bulgularından GGM modelinin verilerin normallik varsayımı dışında başarılı olmaması sebebiyle alternatif modeller incelenmiş ve yüksek bağımlı verilerde tercih edilen, parametrik olmayan ve lineer regresyona dayalı olan genelleştirilmiş toplamalı modellerin (generalized additive models) uygun olduğu görülmüştür. Nitekim bu modeller biyolojik sistem uygulamalarında uygulanırken Denklem (1)'deki ifadenin elde edilmesi için orijinal R kodları tekrar düzenlenmiştir. Proje yürütücüsünün kaleme aldığı bu modeller, rassal orman (random forest – RF) ve sınıflandırma ve regresyon ağacı (classification and regression tree), çok değişkenli adaptif regresyon splinleri (multivariate adaptive regression splines – MARS), konik MARS (CMARS) ve dirençli konik MARS (robust conic MARS – RCMARS) olarak sıralanabilir. ODTÜ BAP projesi (proje no: BAP-01-09-2016-002) olarak kabul edilen bu proje 1/1/2016 ve 31/6/2018 tarihleri arasında tamamlanacaktır. Dolayısıyla tamamlanan bu proje içerisinde RF ve CART analiz sonuçları tamamlanmış, her ikisinin de beklendiği gibi özellikle normal olmayan verilerde daha iyi sonuç verdiği, normallik altında ise GGM'e yakın doğruluğa sahip olduğu bulunmuştur. Bu sonuçların elde edilmesinde Gaussian kopula, hem log-normal, student-t ve üssel marjinaline sahip gen verilerin elde edilmesinde hem de yarı normal, yarı üssel ve yarı normal yarı student-t karma dağılımlı verilerin simüle edilmesinde kullanılmıştır. Çalışmaların tamamında ayrıca diğer Archimedian (Nelsen, 2006; Genest ve Fabre, 2007) kopulalar da incelenmiş, ancak student-t ve Gaussian dışındakilerin sadece çok küçük (6 veya 7) değişken sayısı durumunda açık haliyle yazılabildiği, diğer durumlarda tek bir yazılım ve kapalı hallerinin bulunmadığı saptanmıştır. Diğer yandan student-t ve Gaussian kopulalarının yaklaşık aynı sonucu vermesine rağmen, student-t kopulasının işlemsel zorluğundan dolayı, Gaussian kopula, hem açık ifadesi hem de işlemsel basitliği sebebiyle tüm simülasyonlarda tercih edilmiştir. Proje boyunca elde edilen farklı

model sonuçlarının bir kısmı uluslararası konferanslarda tam metin olarak sunulmuş, Springer yayınevinde değerlendirilmesi için bir bölüm yazılmış ve son olarak RF ve GGM kıyaslamasına yönelik bir makale çalışması SCI indeksli bir dergide değerlendirilmesi için gönderilmiştir. İlgili çalışmalara ait bilgiler aşağıda sıralanmaktadır. **Bu bulguların elde edilmesinde ise proje yürütücüsü ve yüksek lisans bursiyeri Deniz Seçilmiş beraber çalışmışlardır.**

(a3) Purutçuoğlu, V. 2015. "Applications of copulas in deterministic models of biochemical networks", Proceedings of Second International Conference on Recent Advances in Pure and Applied Mathematics (ICRAPAM 2015).

(b3) Seçilmiş, D., Purutçuoğlu, V. 2016. "Random forest algorithms in the construction of biological networks", (hakem değerlendirmesinde), 1-21.

(c3) Seçilmiş, D, Purutçuoğlu, V. 2017. "Nonparametric versus parametric models in inference of protein-protein interaction networks", Proceeding of International Conference on Advances in Science and Arts Istanbul 2017 (ICASA 2017), 55-61.

(d3) Seçilmiş, D., Purutçuoğlu, V. 2016. "Copulas in random forest algorithm for modelling biochemical networks", (hakem değerlendirmesinde – kitap bölümü, **ilk değerlendirme süreci tamamlanarak, yeni hakem düzeltmelerinin ilave edilmesi için çalışılmaktadır**), 1-12.

Bunun yanı sıra projede bursiyer olarak görev alan yüksek lisans öğrencisi Deniz Seçilmiş tezinin yazımını tamamlayarak Nisan 2017'de tezini savunmuş ve başarılı bulunarak yüksek lisans programını tamamlamıştır. Doktora öğrencisi Hajar Farnoudkia ise CGGM ve RJMCMC ve Gibbs örnekleme dayalı çalışmalarını toparlayarak, ders aşamasında olduğu doktora programında bir yandan tezinin ilk bölümlerini yazmaya başlamıştır. Proje bitmesine rağmen zaman serisi CGGM modeli üzerine çalışmaya ise proje yürütücüsü ile beraber devam etmektedir.

Son olarak proje boyunca yazılan tüm kodların R programı ile fonksiyonları yazılmış ve ortak kullanıma hazır hale getirilmiştir. Nitekim bu kodların ve tüm proje çıktılarının paylaşılması için bir internet sayfası hazırlanması için proje önerisinde yazılan hizmet kaleminden destek alınmış, bu amaçla avans talebinde de bulunularak, bir proje internet sayfası hazırlığı yapılmıştır. Ancak daha sonra faturalandırmada firmanın tekrifatlı fatura verme konusundaki isteksizliği sebebiyle uygun fatura alınamadığından ödeme yapılamamış ve avans olarak çekilen bütçe, proje hesabına tekrar yatırılmıştır. **Bu sebeple proje web sayfası hazırlanması için ayrılan bütçe hiç kullanılmadan proje hesabında kalmıştır.** Ancak ilerleyen zamanlarda statik bir proje sayfasının açılabilmesi için ODTÜ Bilişim Hizmetleri

ofisine başvurulmuş ve gerekli işlemlere başlanılmıştır. **Diğer yandan “Bulgular” Bölümünde listelenen tüm kabul edilmiş, yayınlanmış ve hakem sürecinde olan çalışmaların kopyaları, “Bulgular” bölümünde verilen sıralama ile (a1, b1, c1, d1, a2, a3, b3, c3, ve d3) proje ekleri olarak projenin Grup Yürütme Kurulu’na sunulmuştur.**

5. Tartışma ve Sonuç

Bu projede ilk defa CGGM modeli biyolojik sistem yapılarını açıklamada kullanılmış, RJMCMC algoritması ise yüksek boyutlu ve çok değişkenli sistem yapılarının model parametre tahminini yapmak için orijinal R kodları yazılarak uygulanmıştır. Çalışmanın modellemeye dayalı bu ilk aşamasında CGGM modelinin mevcut GGM modeline göre daha iyi sonuç verdiği bulunmuştur. Diğer yandan sistem biyolojisi literatürde daha yeni önerilen ve RJMCMC algoritmasına alternatif olarak düşünülen BDMCMC algoritması sonuçlarına göre de bu projede uygulanan RJCMCM tahmin sonuçlarının yüksek doğruluk oranına sahip oldukları görülmüştür. Bu da RJMCMC ile tahmin edilen CGGM'in, karmaşık biyolojik sistem yapılarını matematiksel yollarla açıklamada başarılı olduğunu göstermektedir. Diğer yandan proje önerisinde önerilmemesine rağmen RJMCMC'yi hızlandırmaya yönelik literatür taraması yapılmış ve seçilen Gibbs örneklem metodunun buna iyi bir alternatif olabileceği gözlemlenmiştir. Yapılan RJMCMC ve Gibbs örneklem karşılaştırma sonuçlarından da ikisinin tamamen aynı doğruluğa sahip olmasına rağmen, Gibbs örneklem metodunun bazı sistem yapılarında hesaplama süresini kısalttığı bulunmuştur. Önerilen CGGM ifadesini iyileştirmeye yönelik olarak da aslında bu modelin özellikle zamana bağlı olmayan verilerde tercih edildiği bulunmuş ve modele zaman serisi ifadesinin ilave edilmesi halinde nasıl yazılabileceği ve hangi tahmin metodu ile model parametrelerinin elde edilebileceği araştırılmıştır. Bu araştırmalar sonunda zamana bağlı CGGM denklemin yaratıldığı bulunmuş, ancak bu modele uygun fazla model parametre tahmin metodu geliştirme çalışması yapılmadığı görülmüştür. Bu konuda yapılan araştırmalar sonunda aslında bu geniş model için de bir Bayesci olan ve olmayan iki ayrı parametre tahmin yolunun uygulanabileceği saptanmıştır. Bu model ve önerilen tahmin metodunun ise, bu sonuç raporunun kabul edilmesi halinde, projenin yeni araştırma konusu olarak TÜBİTAK 1002 proje önerisi biçiminde sunulmasına karar verilmiş ve ilgili proje önerisinin yazımına başlanmıştır.

Çalışmanın ikinci aşamasını kapsayan ve GGM' nin detaylı performans değerlendirmesinde ise GGM ifadesinin aslında oldukça sınırlı bir veri seti için başarılı olduğu gözlemlenmiştir. Verinin normallikten uzaklaşması halinde doğruluk değeri belirgin derecede düşmekte, hatta

sıklıkla hesaplanamamaktadır. Hesaplanamama nedeni ise modelin aslında seyrek olan ve normal olmayan sistem verilerinde hiç bir ilişkiyi bulamaması, dolayısıyla aslında var olan tüm bağlantıları sıfır olarak hesaplaması sebebiyle ilgili doğruluk ölçüsünde tanımsız ifadelerin elde edilmesi, bu ölçüye dayalı F-ölçeği ve MCC gibi karma ölçüleri de hesaplayamamasından kaynaklanmaktadır. Bu sebeple normal dağılımdan uzaklaşılması halinde yeni modelleme yollarının geliştirilmesi gündeme gelmiş ve proje yürütücüsü bu projenin devam ettiği sırada ODTÜ BAP projesi olarak yeni bir proje önerisinde bulunmuştur. Bu yeni projede RF metodu ve CART metotlarının GGM'e bu durumlarda iyi bir alternatif olabileceği önerilmiştir. Bu konuda proje yürütücüsü tarafından önerilen tek metot RF ve CART olmamakla beraber, bu projenin tamamlanması sebebiyle, önerilen diğer metotlar (CMARS ve RCMARS) halen devam eden bu BAP projesinin çıktıları olarak sunulacaktır. Fakat bu proje ve devam eden BAP projesi göstermiştir ki biyolojik sistemlerin ifade edilmesinde parametrik alternatifler içinde CGGM modeli GGM'den daha başarılı sistem yapılarının açıklayabilse de, özellikle parametrik olmayan yeni modellerin geliştirilmesine ve normallik kısıtının kaldırılabilceği yeni modellerin yaratılmasına ihtiyaç duyulmaktadır. Diğer yandan tüm bu modelleme çalışmalarında, Gaussian kopulanın farklı biyolojik sistem verileri yaratmada en etkili yöntem olduğu gözlemlenmiştir.

Bunun yanı sıra bu proje ile birlikte bir doktora ve bir yüksek lisans öğrencisinin sistem yapılarının modelleme ve parametre tahmin yöntemleri üzerine yetiştirilmesi sağlanmış, yüksek lisans öğrencisinin bu konudaki tezi başarı ile savunulmuştur. Ayrıca tüm proje boyunca elde edilen sonuçlar uluslararası konferanslarda sözlü sunum olarak paylaşılarak, elde edilen sonuçların düzenli biçimde bilimsel ortamlarda tartışılması sağlanmıştır. Nitekim bu sunumların tamamı SCI indeksli makale, Springer yayınevi tarafından değerlendirilecek olan kitap bölümleri ve uluslararası konferans tam metin bildirimleri olarak da kaleme alınmış ve bunlardan tam metin bildirimlerin ve özet bildirimlerin yayını kabul edilmiştir. Makale ve kitap bölümlerinin hakemlik süreci ise devam etmekte, ancak yakın zamanda öncelikle kitap bölümlerine ait sonuçların tamamlanması beklenmektedir.

Sonuç olarak proje bulgularının hem karmaşık sistemleri daha iyi ifade etmede iyi sonuçlar verdiği, önerilen metotların başka sistem hastalıklarını ve yolak yapılarını açıklamada kullanılabileceği bulunmuştur. Ayrıca tamamlanan proje hedeflerine ulaşılmış ve proje önerisinden bağımsız yeni alt başlıklar da araştırmaya dahil edilmiştir. Böylece projenin bilimsel literatüre, yetiştirilen öğrencilerle de topluma katkı yapması sağlanmıştır.

Kaynaklar

- Barabási, A. L., Oltvai, Z. N. 2004. "Network biology: understanding the cells functional organization", *Nature Reviews Genetics*, 5, 101-113.
- Bolouri, H. 2008. *Computational Modeling of Gene Regulatory Networks*. Imperial College Press.
- Bower, J. M., Bolouri, H. 2001. *Computational Modelling of Genetic and Biochemical Networks (2nd Edition)*. Massachusetts Institute of Technology.
- Boys, R. J., Wilkinson, D., Kirkwood, T. B. L. 2008. "Bayesian inference for a discretely observed stochastic kinetic model", *Statistical Computation*, 18, 125-135.
- Carlin, B. P., Chib, S. 1995. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of Royal Statistical Society, Series B*, 57 (3), 473-484.
- Defferli, Ö, Purutçuoğlu, V., Weber, G. W. 2014. "Advance mathematical and statistical tools in the dynamic modelling and simulation of gene-environment networks". *Modeling, Optimization, Dynamics and bioeconomy*. Editor: Zilberman, D., Pinto, A. Springer-Verlag
- Dobra, A., Lenkoski, A. 2010. "Copula Gaussian graphical models and their application to modeling functional disability data", *Annals of Applied Statistics*, 5 (2A), 969-993.
- Endy, D., Brent, R. 2001. "Modelling cellular behaviour", *Nature*, 409, 391-395.
- Meinshausen, N., Bühlmann, P. 2006. "High-dimensional graphs and variable selection with the lasso", *The Annals of Statistics*, 34 (3), 1436-1462.
- Friedman J, Hastie T, Tibshirani R. 2008. "Sparse inverse covariance estimation with the graphical lasso", *Biostatistics*, 9, 432- 441.
- Gamerman, D., Lopes, H. F. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd Edition)*. Chapman and Hall/CRC.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. 2004. *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Genest, C., Favre, A.-C. 2007. "Everything you always wanted to know about copula modeling but were afraid to ask", *Journal of Hydrologic Engineering*, 12, Special Issue: Copulas in Hydrology, 347-368.

Gilks, W., Richardson, S., Spiegelhalter, D. 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC.

Gillespie, D. T. 1977. "Exact stochastic simulation of coupled chemical reactions", *Journal of Physical Chemistry*, 81 (25), 2340-2361.

Golightly, A., Wilkinson, D. 2011. "Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo", *Interface Focus*, 1, 807-820.

Green, P. J. 1995. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, 82 (4), 711-732.

He, Y., Zhang, X., Wang, P., Zhang, L. 2017. "High dimensional Gaussian copula graphical model with FDR control", *Computational Statistics and Data Analysis*, 113, 457-474.

Hsieh, C. J., Sustik, M. A., Dhillon, I.S., Ravikumar, P. 2014. "QUIC: Quadratic Approximation for Sparse Inverse Covariance Estimation", *Journal of Machine Learning Research*, 15, 2911-2947.

Jong, H. D. 2002. "Modelling and simulation of genetic regulatory systems: a literature review", *Journal of Computational Biology*, 9 (1), 67-103.

Junker, B. H., Schreiber, F. H. 2008. *Analysis of Biological Networks*. John Wiley and Sons.

Khanin, R., Wit, E. 2006. "How scale-free are gene networks?", *Journal of Computational Biology*, 13 (3), 810-818.

Kolch, W., Calder, M., Gilbert, D. 2005. "When kinases meet mathematics: the systems biology of MAPK signalling", *FEBS Letters*, 579, 1891-1895.

Li, H. 2008. "Statistical methods for inference of genetic networks and regulatory modules". *Analysis of Microarray Data, a Network-based Approach*. Editors: Emmert-Streib, E., Dehmer, M. Weinheim: Wiley-VCH Verlag.

Liu, H., Han, F., Yuan, M., Lafferty, J., Wasserman, L. 2012. "High dimensional semiparametric gaussian copula graphical models", *The Annals of Statistics*, 40 (4), 2293-2326

Liu, W., 2013. "Gaussian graphical model estimation with false discovery rate control", *The Annals of Statistics*, 41 (6), 2948 - 2978.

Liu, H., Roeder, K., Wasserman, L. 2010. "Stability approach to regularization selection (stars) for high dimensional graphical models", Proceedings of the 23rd International Conference on Neural Information Processing Systems, 1432–1440.

Maiwald, T., Schneider, A., Busch, H., Sahle, S., Gretz, N., Weiss, T., Kummer, U., Klingmuller, U. 2010. "Combining theoretical analysis and experimental data generation reveals IRF9 as a crucial factor for accelerating interferon induced early antiviral signalling", FEBS Journal, 277, 4741-4754.

Mohammadi, A., Wit, E. 2015. "BDgraph: Bayesian structure learning of graphs in R", Bayesian Analysis, 10, 109-138.

Nelsen, R. B. 2006. An Introduction to Copulas. Springer Science-Business Media, Inc.

O'Hagan, A., Forster, J. J. 2004. Kendall's Advanced Theory of Statistics. Edward Arnold Publishing.

O'Shea, J. J., Gadina, M., Schreiber, R. D. 2002. "Cytokine signalling in 2002: New surprises in the JAK/STAT pathway", Cell, 109: S121-S131.

Purutçuoğlu, V., Wit, E., 2008. "Bayesian inference for the MAPK/ERK pathway by considering the dependency of the kinetic parameters", Bayesian Analysis, 3 (4), 851-886.

Purutçuoğlu, V. 2012. "Inference of the stochastic MAPK pathway by modified diffusion bridge method", Central European Journal of Operations Research, doi: 10.1007/s10100-012-0237-8, 1-15.

Purutçuoğlu, V. , Farnoudkia, H. 2017. "Gibbs sampling in inference of the copula Gaussian graphical model adapted to biological networks", Acta Physica Polonica Seres A, 1-12 (Kabul edildi).

Purutçuoğlu, V., Ayyıldız, E., Wit, E. 2017. Comparison of two inference approaches in Gaussian graphical models. Turkish Journal of Biochemistry, 1-9.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. and Nolan, G. 2005. "Causal protein signaling networks derived from multiparameter single-cell data, Science, 308 (5721), 523–529.

Stranger, B., Nica, A., Forrest, M., Dimas, A., Bird, C., Beazley, C., Ingle, C., Dunning, M., Flicek, P., Montgomery, S., Tavare, S., Deloukas, P., Dermitzakis, E. 2007. "Population genomics of human gene expression", Nature Genetics 39, 1217–1224.

- Tibshirani, R. 1996. "Regression shrinkage and selection via the lasso", *Journal of Royal Statistical Society-B*, 58 (1), 267-288.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K. 2005. "Sparsity and smoothness via fused lasso", *Journal of the Royal Statistical Society*, 67 (1), 91-108.
- Trivedi, P., Zimmer, D. M. 2005. "Copula modeling: an introduction for practitioners", *Foundations and Trends in Econometrics*, 1 (1), 1-111.
- Van Kampen, N. G. 1981. *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland.
- Vyshechirsky, V., Girolami, M., Gormand, A., Kolch, W. 2006. "A bayesian analysis of the ERK signalling pathway", DCS Technical Report Series TR-2006-227, Department of Computing Science, University of Glasgow.
- Walker, S. 2009. "A Gibbs sampling alternative to reversible jump MCMC", *Electronic Journal of Statistics*, 1-3.
- Wawrzyniak, M. M., Kurowicka, D. 2006. *Dependence concepts*. Delft University of Technology, Delft Institute of Applied Mathematics, Delft, Netherlands.
- Wang, T., Ren, Z., Ding, Y., Fang, Z., Sun, Z., MacDonald, M. L., Sweet, R. A., Wang, J., Chen, W. 2016. "FastGGM: An efficient algorithm for the inference of Gaussian graphical model in biological networks", *PLOS Computational Biology*, 12 (2), 1-16.
- Wilkinson, D. J. 2006. *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC.
- Whittaker, J. 1990. *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, New York.
- Xu, D., Yin, C., Wang, S., Xiao, Y. 2013. "JAK-STAT in lipid metabolism of adipocytes", *JAKSTAT*. 2(4), e27203.
- Zhou, S. and Shen, X. 2012. "Spatially adaptive regression splines and accurate knot selection schemes", *Journal of American Statistical Association*, 96, 247–259.

Zou, H. 2006. "The adaptive lasso and its oracle properties", *Journal of the American Statistical Association*, 101, 1418-1429.

Zou, H., Hastie, T. 2005. "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society*, 67 (2), 301-320.

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje Yürütücüsü:	Doç. Dr. VİLDA PURUÇUOĞLU GAZİ
Proje No:	114E636
Proje Başlığı:	Biyolojik Ağların Gaussian Grafiksel Modellerle Tahmininde Kopulaların Uygulanması Ve Parametre Tahmini
Proje Türü:	Uluslararası
Proje Süresi:	24
Araştırmacılar:	
Danışmanlar:	TAMAY ŞEKER, ALEXANDER DE LEON (Yurt Dışı), ERNST WIT (Yurt Dışı)
Projenin Yürütüldüğü Kuruluş ve Adresi:	ORTA DOĞU TEKNİK Ü. FEN-EDEBİYAT F. İSTATİSTİK B.
Projenin Başlangıç ve Bitiş Tarihleri:	01/04/2015 - 01/04/2017
Onaylanan Bütçe:	171828.0
Harcanan Bütçe:	120720.38
Öz:	Proje gerçekçi büyüklükteki karmaşık biyolojik ağ yapılarının sistem davranışlarını durağan haldeyken modellenmesi ve model parametrelerinin Bayesci yaklaşımlarla tahmin edilmesini kapsamaktadır. Modellemede kopula Gaussian grafiksel modeli (CGGM) kullanılmış ve parametre tahmini ise öncelikli olarak ters atlamalı Markov zinciri Monte Carlo yöntemiyle yapılmıştır.
Anahtar Kelimeler:	Kopula Gaussian grafiksel modeli, ters atlamalı Markov zinciri Monte Carlo algoritması, kopula
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Hayır
Projeden Yapılan Yayınlar:	1- Modelling of biological networks via copula Gaussian graphical model and bayesian inference of model parameters (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 2- Bayesian inference of deterministic MAPK/ERK pathway via reversible jumps Monte Carlo method (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 3- Copulas in Random Forest Algorithm for Modelling Biochemical Networks (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 4- Inference of biological networks via random forest algorithm (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 5- Reconstruction of biological networks via MARS model based on various precision matrices (Bildiri - Uluslararası Bildiri - Sözlü Sunum), 6- Applications of copulas in deterministic models of biochemical networks (Bildiri - Uluslararası Bildiri - Sözlü Sunum),