Bioinformatics: Network Analysis

Networks as a Guiding Tool

COMP 572 (BIOS 572 / BIOE 564) - Fall 2013
Luay Nakhleh, Rice University
Networks have been used to

- guide GWAS
- predict protein function
- model epidemics
- ...

Efficient network-guided multi-locus association mapping with graph cuts
Chloé-Agathe Azencott¹,*\textsuperscript{,}, Dominik Grimm¹, Mahito Sugiyama¹, Yoshinobu Kawahara² and Karsten M. Borgwardt¹,³

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{networks.png}
\caption{Small examples of the three types of networks considered}
\end{figure}

(\textbf{a})
\begin{itemize}
\item Genomic sequence network: SNPs adjacent on the genomic sequence are connected to each other.
\end{itemize}

(\textbf{b})
\begin{itemize}
\item Gene membership network: In addition, SNPs near the same gene (within a specified distance) are connected.
\end{itemize}

(\textbf{c})
\begin{itemize}
\item Gene-interaction network: In addition, SNPs near two interacting genes are connected.
\end{itemize}
Network-Assisted Investigation of Combined Causal Signals from Genome-Wide Association Studies in Schizophrenia

Peilin Jia¹,², Lily Wang³, Ayman H. Fanous⁴,⁵,⁶,⁷, Carlos N. Pato⁷, Todd L. Edwards⁸,⁹, The International Schizophrenia Consortium¹⁰, Zhongming Zhao¹,²,¹⁰*

**Abstract**

With the recent success of genome-wide association studies (GWAS), a wealth of association data has been accomplished for more than 200 complex diseases/traits, proposing a strong demand for data integration and interpretation. A combinatorial analysis of multiple GWAS datasets, or an integrative analysis of GWAS data and other high-throughput data, has been particularly promising. In this study, we proposed an integrative analysis framework of multiple GWAS datasets by overlaying association signals onto the protein-protein interaction network, and demonstrated it using schizophrenia has been particularly promising. In this study, we proposed an integrative analysis framework of multiple GWAS datasets by overlaying association signals onto the protein-protein interaction network, and demonstrated it using schizophrenia.

**Introduction**

Genome-wide association (GWA) studies have, during the past decade, become a powerful tool to study the genetic components of complex diseases [1]. Although an increasing number of genes/variants have been uncovered in GWA studies, which have provided us important insights into the underlying genetic basis of the diseases. Most of these variants have been missed in single marker analysis, as investigators typically employ a genome-wide significance cutoff.
Network-based prediction of protein function
Roded Sharan¹, Igor Ulitsky¹ and Ron Shamir*
Epidemics and Networks
Diseases and the Networks that Transmit Them

- The patterns by which epidemics spread through a population is determined not just by the properties of the pathogen carrying it (contagiousness, the length of the infection period, severity, etc.), but also by network structures within the population it is affecting.

- The opportunities for a disease to spread are given by a contact network: there is a node for each person, and an edge if two people come into contact with each other in a way that makes it possible for the disease to spread from one to the other.
Diseases and the Networks that Transmit Them

• Accurately modeling the underlying network is crucial to understanding the spread of an epidemic.
...how travel patterns within a city affect the spread of disease

Modelling disease outbreaks in realistic urban social networks

Stephen Eubank¹, Hasan Guclu², V. S. Anil Kumar¹, Madhav V. Marathe¹, Aravind Srinivasan³, Zoltán Toroczkai⁴ & Nan Wang⁵

Network theory and SARS: predicting outbreak diversity

Lauren Ancel Meyers¹,², *,¹, Babak Pourbohloul³,¹,², M.E.J. Newman²,⁴, Danuta M. Skowronski³,², Robert C. Brunham³,²
...how travel patterns via the worldwide airline network affect the spread of disease

The role of the airline transportation network in the prediction and predictability of global epidemics

Vittoria Colizza*, Alain Barrat†, Marc Barthélemy‡, and Alessandro Vespignani§
Contact networks are also important in understanding how diseases spread through animal populations (e.g., the 2001 foot-and-mouth outbreak in the UK) and plant populations.

Similar models have been employed for studying the spread of computer viruses...
Diseases and the Networks that Transmit Them

- The pathogen and the network are closely intertwined: even within the same population, the contact networks for two different diseases can have very different structures, depending on the diseases’ respective modes of transmission.

- (Think of airborne transmission based on coughs and sneezes, compared to a sexually transmitted disease, and think of the density of the contact networks!)
Branching Processes

- The simplest model of contagion: every person is in contact with k people

- First wave: a person carrying a new disease enters a population and transmits it to each of his contacts independently with probability p.

- Second wave: each person in the first wave transmits to each of his contacts independently with probability p (the contacts of people are mutually exclusive)

- and so on..
Branching Processes

(a) The contact network for a branching process
Figure 21.1: The branching process model is a simple framework for reasoning about the spread of an epidemic as one varies both the amount of contact among individuals and the level of contagion.

(b) With high contagion probability, the infection spreads widely
Branching Processes

(a) The contact network for a branching process

(b) With high contagion probability, the infection spreads widely

(c) With low contagion probability, the infection is likely to die out quickly

Figure 21.1: The branching process model is a simple framework for reasoning about the spread of an epidemic as one varies both the amount of contact among individuals and the level of contagion.
The Basic Reproductive Number $R_0$

- The **basic reproductive number**, denote $R_0$, is the expected number of new cases of the disease caused by a single individual.

- For the simple branching process we saw, we have $R_0 = kp$. 
The Basic Reproductive Number $R_0$

Claim: If $R_0 < 1$, then with probability 1, the disease dies out after a finite number of waves. If $R_0 > 1$, then with probability greater than 0 the disease persists by infecting at least one person in each wave.
The Basic Reproductive Number $R_0$

- Implication: It’s always good to reduce the value of $R_0$!
- Quarantining people reduces $k$ and encouraging behavioral measures such as sanitary practices reduces $p$. 
Clearly, the branching process is too simplistic!
The SIR Epidemic Model

- An individual node goes three potential stages during the course of the epidemic:
  - **Susceptible**: Before the node has caught the disease, it is susceptible to infection from its neighbors.
  - **Infectious**: Once the node has caught the disease, it is infectious and has some probability of infecting each of its susceptible neighbors.
  - **Removed**: After a particular node has experienced the full infectious period, this node is removed from consideration.
The SIR Epidemic Model

- Given this three-stage “life cycle” for the disease at each node, a model for epidemics on networks can be defined.

- The network structure: a directed graph representing the contact network (edge u to v means that if u becomes infected, the disease has the potential to spread to v).

- Two other quantities: p (the probability of contagion) and t_I (the length of infection)
The SIR Epidemic Model

- Initially, some nodes are in the $I$ state and all others are in the $S$ state.
- Each node $v$ that enters the $I$ state remains infectious for a fixed number of steps $t_I$.
- During each of these $t_I$ steps, $v$ has a probability $p$ of passing the disease to each of its susceptible neighbors.
- After $t_I$ steps, node $v$ is no longer infectious or susceptible to further bouts of the disease; we describe it as removed ($R$), since it is now an inert node in the contact network that can no longer either catch or transmit the disease.
The SIR Epidemic Model

Figure 21.2: The course of an SIR epidemic in which each node remains infectious for a number of steps equal to $t_I = 1$. Starting with nodes $y$ and $z$ initially infected, the epidemic spreads to some but not all of the remaining nodes. In each step, shaded nodes with dark borders are in the Infectious (I) state and shaded nodes with thin borders are in the Removed (R) state.

Extensions to the SIR model. Although the contact network in the general SIR model can be arbitrarily complex, the disease dynamics are still being modeled in a simple way. Contagion probabilities are set to a uniform value $p$, and contagiousness has a kind of "on-off" property: a node is equally contagious for each of the $t_I$ steps while it has the disease. However, it is not difficult to extend the model to handle more complex assumptions. First, we can easily capture the idea that contagion is more likely between certain pairs of nodes by assigning a separate probability $p_{v,w}$ to each pair of nodes $v$ and $w$ for which $v$ links to $w$ in the directed contact network. Here, higher values of $p_{v,w}$ correspond to closer contact and more likely contagion, while lower values indicate less intensive contact. We can also choose to model the infectious period as random in length, by assuming that an infected node has a probability $q$ of recovering in each step while it is infected, while leaving
The SIS Epidemic Model

- The SIR epidemic model is appropriate for epidemics in which each individual contracts the disease at most once.

- To allow for nodes that can be reinfected multiple times, a model can have only the S and I, but not R, states.
The SIS Epidemic Model

- Initially, some nodes are in the $I$ state and all others are in the $S$ state.
- Each node $v$ that enters the $I$ state remains infectious for a fixed number of steps $t_I$.
- During each of these $t_I$ steps, $v$ has a probability $p$ of passing the disease to each of its susceptible neighbors.
- After $t_I$ steps, node $v$ is no longer infectious, and it returns to the $S$ state.
The SIS Epidemic Model

In each step, the nodes in the Infectious state are shaded.

21.4 The SIS Epidemic Model

In the previous sections we have been considering models for epidemics in which each individual contracts the disease at most once. However, a simple variation on these models allows us to reason about epidemics where nodes can be reinfected multiple times.

To represent such epidemics, we have nodes that simply alternate between two possible states: Susceptible (S) and Infectious (I). There is no Removed state here; rather, after a node is done with the Infectious state, it cycles back to the Susceptible state and is ready to catch the disease again. Because of this alternation between the S and I states, we refer to the model as the SIS model.

Aside from the lack of an R state, the mechanics of the model follow the SIR process very closely.

• Initially, some nodes are in the I state and all others are in the S state.
• Each node $v$ that enters the I state remains infectious for a fixed number of steps $t_I$.
• During each of these $t_I$ steps, $v$ has a probability $p$ of passing the disease to each of its susceptible neighbors.
• After $t_I$ steps, node $v$ is no longer infectious, and it returns to the S state.

Figure 21.5 shows an example of the SIS model unfolding on a three-node contact network with $t_I = 1$. Notice how node $v$ starts out infected, recovers, and later becomes infected again — we can imagine this as the contact network within a three-person apartment, or a three-person family, where people pass a disease on to others they’re living with, and then get it back from them later.

As with the SIR model, the SIS model can be extended to handle more general kinds of assumptions: different contagion probabilities between different pairs of people; probabilistic...
A Connection Between SIR and SIS

To represent the SIS epidemic using the SIR model, we use a “time-expanded” contact network. The SIS epidemic can then be represented as an SIR epidemic on this time-expanded network.

(a) To represent the SIS epidemic using the SIR model, we use a “time-expanded” contact network.

(b) The SIS epidemic can then be represented as an SIR epidemic on this time-expanded network.
The SIRS Epidemic Model

* Combines SIR and SIS:
  * After an infected node recovers, it passes briefly through the R state on its way back to the S state.
Small World Effect in an Epidemiological Model

Marcelo Kuperman$^{1,*}$ and Guillermo Abramson$^{1,2,†}$

Recall: the small-world model ($p$: probability of rewiring an edge)

The number of infected people ($n_{\text{inf}}(t)$) by SIRS epidemic:

FIG. 1. Fraction of infected elements as a function of time. Three time series are shown, corresponding to different values of the disorder parameter $p$, as shown in the legends. Other parameters are $N = 10^4$, $K = 3$, $\tau_I = 4$, $\tau_R = 9$, $n_{\text{inf}}(0) = 0.1$. 
Acknowledgments

- Slides on epidemics and networks are based on the book “Networks, Crowds, and Markets” by Easley and Kleinberg.