

# Network Science

<http://journals.cambridge.org/NWS>

Additional services for **Network Science**:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



---

## Networks' characteristics are important for systems biology

ANDREW K. RIDER, TIJANA MILENKOVIĆ, GEOFFREY H. SIWO, RICHARD S. PINAPATI, SCOTT J. EMRICH, MICHAEL T. FERDIG and NITESH V. CHAWLA

Network Science / Volume 2 / Issue 02 / August 2014, pp 139 - 161

DOI: 10.1017/nws.2014.13, Published online: 03 September 2014

**Link to this article:** [http://journals.cambridge.org/abstract\\_S2050124214000137](http://journals.cambridge.org/abstract_S2050124214000137)

### How to cite this article:

ANDREW K. RIDER, TIJANA MILENKOVIĆ, GEOFFREY H. SIWO, RICHARD S. PINAPATI, SCOTT J. EMRICH, MICHAEL T. FERDIG and NITESH V. CHAWLA (2014). Networks' characteristics are important for systems biology. Network Science, 2, pp 139-161 doi:10.1017/nws.2014.13

**Request Permissions :** [Click here](#)

# *Networks' characteristics are important for systems biology*

ANDREW K. RIDER and TIJANA MILENKOVIĆ

*Department of Computer Science and Engineering, ECK Institute for Global Health, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN 46556, USA*

GEOFFREY H. SIWO and RICHARD S. PINAPATI

*Department of Biological Sciences, ECK Institute for Global Health, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN 46556, USA*

SCOTT J. EMRICH

*Department of Computer Science and Engineering, ECK Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA*

MICHAEL T. FERDIG

*Department of Biological Sciences, ECK Institute for Global Health, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN 46556, USA*

NITESH V. CHAWLA

*Department of Computer Science and Engineering, ECK Institute for Global Health, Interdisciplinary Center for Network Science and Applications (iCeNSA), University of Notre Dame, Notre Dame, IN 46556, USA  
(e-mail: nchawla@nd.edu)*

---

## **Abstract**

A fundamental goal of systems biology is to create models that describe relationships between biological components. Networks are an increasingly popular approach to this problem. However, a scientist interested in modeling biological (e.g., gene expression) data as a network is quickly confounded by the fundamental problem: how to construct the network? It is fairly easy to construct a network, but is it *the* network for the problem being considered? This is an important problem with three fundamental issues: How to weight edges in the network in order to capture actual biological interactions? What is the effect of the type of biological experiment used to collect the data from which the network is constructed? How to prune the weighted edges (or what cut-off to apply)? Differences in the construction of networks could lead to different biological interpretations.

Indeed, we find that there are statistically significant dissimilarities in the functional content and topology between gene co-expression networks constructed using different edge weighting methods, data types, and edge cut-offs. We show that different types of known interactions, such as those found through Affinity Capture-Luminescence or Synthetic Lethality experiments, appear in significantly varying amounts in networks constructed in different ways. Hence, we demonstrate that different biological questions may be answered by the different networks. Consequently, we posit that the approach taken to build a network can be matched to biological questions to get targeted answers. More study is required to

understand the implications of different network inference approaches and to draw reliable conclusions from networks used in the field of systems biology.

---

**Keywords:** *systems biology, network inference*

## 1 Introduction

### 1.1 Motivation and background

High-throughput biological data such as protein–protein interactions (PPIs), gene expression profiles, and metabolic interactions contain information about how different components of a cell interact in concert and can be used, for example, to elucidate potential drug targets and to further our understanding of disease (van't Veer et al., 2002; van de Vijver et al., 2002). Because biological data are generally noisy and expensive to obtain, modern “systems biology” has produced integrative analysis frameworks to help overcome the challenges and pitfalls of these important collections.

One particular analysis is network-driven, where genes can be nodes and specific—but different—approaches are often used to infer edges that predict relationships between them (Nayak et al., 2009; Ivliev et al., 2010; Feizi et al., 2013). For example, gene co-expression networks constructed with correlation-based measures have been used to identify transitive relationships (Zhou et al., 2002), gene regulatory patterns (van Noort et al., 2004), and biological modules (Mason et al., 2009). Further, they have been successfully combined with transcription factor, eQTL, and PPI data into integrative Bayesian networks (Zhu et al., 2008). Gene co-expression networks are therefore one commonly used example of a complex model built from the high-throughput biological data collections. For this reason, we focus on gene co-expression networks, but our study is applicable to any type of biological networks.

Despite their importance in systems-level analysis, however, there is currently no consensus about how to construct biological networks and gene co-expression networks in particular. Moreover, the link between a chosen network inference method to network topology and functional content (i.e., how to generate edges that encapsulate some biological knowledge) is not well understood. If the choice of a specific network construction has significant effects on the subsequent interpretation, biological advances will be significantly challenged unless done with considerable care.

To demonstrate this point, we investigate how variables in network construction affect the content of the resulting networks. Of all biological network types, we focus on gene co-expression networks as a proof of concept. Gene co-expression networks are commonly constructed as follows. Given a gene expression dataset, the strength (or weight) of the co-expression relationship between each pair of genes can be quantified in a number of ways, each of which may be revealing different biological insights (Rider et al., 2011; Mason et al., 2009). Then, an arbitrary cut-off is chosen in order to include only the strongest edges into the network (Butte & Kohane, 2000; De Smet & Marchal, 2010; Markowitz & Spang, 2007; Barabasi & Oltvai, 2004). If the edge strength is positively correlated with the underlying biological knowledge, then the choice of a cut-off can be a major factor in determining the network's ability to accurately uncover this knowledge.

Here we report significant differences in the biological information captured using two different types of gene expression datasets and different choices of network inference variables. This work shows that substantial care must accompany network analysis and is driven by the principle of induction: to conclude that a network construction variable has an effect, it is enough to find an example of two different choices for the variable (e.g., two different edge weighting measures) that result in significant differences; this constitutes a proof that the choice for that variable matters. And given one such example, it stands to reason that there may be other choices for the variable (e.g., other edge weighting measures) that also result in network differences. Because many other network inference methods rely on choosing an edge weighting method and a cut-off, and because in addition to gene co-expression networks many other biological network types rely on network inference, this study also has broad implications.

## 1.2 Our approach and contributions

In brief, we studied the effects of different edge weighting methods, data types, and edge cut-offs on the functional content and topology of resulting gene co-expression networks.

### 1.2.1 Edge weighting methods

To maximize the significance of the observations derived from the experiments discussed in the next sections, we chose popular edge weighting methods: correlation-based measures (Pearson's correlation in particular) and mutual information (MI) (see Methods). Pearson's correlation is vulnerable to perturbation from outliers and is constrained in that it captures only linear relationships. Mutual information is the common alternative, and has been used extensively in relevant frameworks such as ARACNE (Margolin et al., 2006), MRNET (Meyer et al., 2008), and CLR (Faith et al., 2007).

### 1.2.2 Data types

Because not all genes are active at all times, under all conditions, or in all strains (Jansen, 2001; Hughes et al., 2000), we studied the effect of the type of biological experiment used to collect gene expression data. We constructed networks from expression data obtained from two types of biological experiments: line cross (Brem & Kruglyak, 2005) and treatment. In the line cross experiment, two strains of yeast (*Saccharomyces cerevisiae*) were crossed and microarray experiments were performed on the strains and their progeny. In the treatment experiment, the same species of yeast was subjected to different chemicals or mutations before having expression levels measured through microarray experiments (Hughes et al., 2000) (see Methods). These are two very different types of experiments, which could affect the functional content and topologies of gene co-expression networks created from the data.

### 1.2.3 Edge cut-offs

For each combination of edge weighting method and data type we first constructed a network containing the top  $k$  strongest edges by varying  $k$  from 0% to 100% of

the strongest edges (in increments of 6,000 edges). For our more detailed analyses, we focused on stronger edges by varying  $k$  from 2,500 to 75,000 of the strongest edges (in increments of 2,500 edges).

### 1.2.4 Evaluation

To evaluate how accurately a given co-expression network captured existing biological knowledge, we tested in a systematic precision-recall setting whether edges in the network corresponded to known interactions (e.g., PPIs), as well as whether genes that were connected by an edge in the network shared Gene Ontology (GO) annotations. This is a common approach to evaluation of biological networks (De Smet & Marchal, 2010; Ucar et al., 2007; Nayak et al., 2009; Ashburner et al., 2000; Faith et al., 2007; Marbach et al., 2010; Grigoriev, 2001). We focused on yeast data types (see above) because yeast is among the most studied and best annotated species to date.

### 1.2.5 Summary

We show that different network inference strategies result in networks that may contain answers to distinct biological questions. These results and conclusions highlight challenges in network construction, along with their impact, which most urgently require the attention of the systems biology community.

## 2 Methods

### 2.1 Data

#### 2.1.1 Gene expression datasets

We use gene expression data resulting from two types of experiments: line cross and treatment.

- In the line cross experiment, expression levels were measured from 130 segregants of a cross of two strains of yeast (Brem & Kruglyak, 2005). Just as Smith & Kruglyak (2008), we used all probes for which more than 80% of the expression data was present, resulting in a total of 5,829 unique open reading frames (ORFs). We measured the strength of relationships between genes by relying on the normalized log ratios provided. Expression values for repeated probes were averaged. Line cross experiments can reveal relationships between heritability and expression (Jansen, 2001).
- In the treatment experiment, 300 expression profiles were generated from mutant or chemical-treated cultures for 6,314 ORFs. We measured the strength of relationships between 6,207 unique ORFs by relying on  $p$ -values resulting from a gene-specific error model that accounts for variation in genes as well as variation across chips (Hughes et al., 2000). Repeated measurements were averaged. Treatment experiments cause perturbations (whether due to environment or mutation) that affect gene expression levels and thus enable one to elucidate relationships between genes that may not be evident when the cell is “at rest” (Pe’er et al., 2001).

### 2.1.2 Known interactions

By known interactions, we mean the set of interactions that are already available in public databases. Known interactions are a valuable collection of ground truth data, but the completeness of the data can vary greatly between organisms. For example, baker's yeast (*Saccharomyces cerevisiae*) is relatively well characterized, whereas non-model organisms, such as the pathogen of malaria (*Plasmodium falciparum*), are not. Hence, in our study, we focus on the well-studied yeast. Known interactions of various types were obtained from SGD (Saccharomyces Genome Database) (Supplementary Table 2). Seventeen of the 26 interaction types have more than 500 genes involved into the corresponding interactions, and 14 of the 26 interaction types have more than 1,000 genes involved into the corresponding interactions (Supplementary Table 2). Hence, known interaction data provides a strong basis for network comparison. Different interaction types describe relationships between genes that were discovered in experiments with different biological meanings. For example, Affinity Capture-mass spectrometry (MS) interactions are determined by using a "bait" protein that is "captured" by a polyclonal antibody or an epitope tag. The associated partners are then identified by MS. Affinity Capture-MS interactions typically correspond to physical interactions between proteins. In contrast, Synthetic Lethality interactions are determined by observing when mutations or deletions in separate genes, each of which causes a minimal change in phenotype alone, result in lethality to a cell.

### 2.1.3 GO data

GO assigns biological process, molecular function, and cellular component labels (i.e., terms) to genes. GO terms are arranged in a hierarchical fashion. Genes that share a GO term are typically functionally-related and may thus be more likely to interact than genes that do not share a GO term. We use GO-slim biological process data (Christie et al., 2009). GO-slim terms are a reduced set of GO terms corresponding to higher levels of the GO hierarchy.

## 2.2 Edge weighting methods

We measure the strength of the relationship between two genes by using either a signed variation of Pearson's correlation or mutual information.

### 2.2.1 Correlation

The signed variation of Pearson's correlation is given in Equation (1) (Mason et al., 2009).

$$\text{signed correlation}(x, y) = \frac{1 + \text{correlation}(x, y)}{2} \quad (1)$$

### 2.2.2 Estimation of mutual information

Mutual information is a measure of the mutual dependence of two random variables (Equation (2)) (Mason et al., 2009).

$$MI(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left[ \frac{p(x, y)}{p(x)p(y)} \right] \quad (2)$$

Calculation of mutual information is complicated by two factors. First, there are a low number of instances from which to estimate probability distributions. Second, expression data are continuous. To overcome these issues, we utilized the Parzen window approach to density estimation (Parzen, 1962). Equation (3) estimates the density function  $\hat{p}(x)$  over  $N$  samples of variable  $x$ :

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N \delta(x - x^{(i)}, h) \quad (3)$$

where  $\delta(\dots)$  is the Parzen window function described in Equation (4),  $x^{(i)}$  is the  $i$ th sample, and  $h$  is the window size. When  $d = 1$ , this equation returns the estimated marginal density. When  $d = 2$ , it gives an estimate of the joint density,  $p(x, y)$ , which can be used to calculate the mutual information in Equation (2).

$$\delta(z, h) = \frac{\exp\left(-\frac{z^T \Sigma^{-1} z}{2h^2}\right)}{(2\pi)^{d/2} h^d |\Sigma|^{(1/2)}} \quad (4)$$

In Equation (4),  $z = x - x^{(i)}$ ,  $h$  is the window size,  $d$  is the dimension of the sample, and  $\sigma$  is the covariance of  $z$ . For further details and the implementation used, see Peng et al. (2005).

### 2.3 Evaluation

Precision and recall are defined in terms of true positives ( $tp$ ), false positives ( $fp$ ), and false negatives ( $fn$ ). In the case of the networks we consider, a  $tp$  is an edge in the network that corresponds to a known interaction or whose end nodes share a GO term. A  $fp$  is an edge in the network that is not among the known interactions or whose end nodes do not share a GO term. A  $fn$  is a pair of genes that are linked by a known interaction or that share a GO term but that are not linked by an edge in the network. Equations (5) and (6) define precision and recall.

$$\text{precision} = \frac{tp}{tp + fp} \quad (5)$$

$$\text{recall} = \frac{tp}{tp + fn} \quad (6)$$

Precision–recall curves have been identified as useful alternatives to receiver operating characteristic (ROC) curves in situations in which there is a large imbalance in the data (Kok & Domingos, 2005; Landgrebe et al., 2006). Namely, with only 220,009 known interactions and 5,892,199 pairs of genes that share a GO term out of approximately 18 million possible edges, the problem of network inference qualifies as imbalanced.

Increase in precision typically results in decrease in recall, and vice versa. The  $F$ -score reconciles precision and recall by combining them into a single score, namely their harmonic mean (Equation (7)).

$$F\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

Given a network constructed using a given edge weighting method, data type, and edge cut-off, for each known interaction type, we counted how many interactions

of the given type are present in the network. Then, we computed  $p$ , the probability of observing the same or higher number of interactions of the same type purely by chance, by using the model of hypergeometric distribution. If we denote by  $N$  the number of possible edges, by  $m$  the total number of known interactions of a given type, by  $n$  the number of edges in the network, and by  $k$  the number of known interactions of a given type that are in the network, the probability of observing exactly  $k$  known interactions in the network purely by chance is computed as shown in Equation (8). Then, to compute probability  $p$ , we sum Equation (8) over all possible numbers of known interactions equal to or greater than  $k$ .

$$P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \quad (8)$$

Given a set of networks corresponding to 30 different edge cut-offs in the [2,500, 75,000] range, where the same edge weighting method and the data type are used for all 30 networks, we form a vector of 30 probabilities  $p$  for the set, where the probabilities are computed for a given known interaction type as explained above. We then compare two vectors of 30 elements corresponding to two network sets constructed using the same data type but different edge weighting methods (Supplementary Table 1). We also compare vectors of 30 elements corresponding to two network sets constructed using the same edge weighting method but different data types (Supplementary Table 1). We do this for each known interaction type. We compare any two vectors by using the Wilcoxon signed-rank test, a nonparametric analog of the  $t$ -test (Demsar, 2006).

#### 2.4 Topological analysis of networks

We use three network topological measures: the clustering coefficient, the closeness centrality, and the betweenness centrality (Milenković et al., 2008; Milenković et al., 2011).

The clustering coefficient of node  $v$  describes the proportion of  $v$ 's neighbors that are connected to each other. It is computed as shown in Equation (9), where  $d$  is the number of neighbors of  $v$  and  $k$  is the number of connected pairs of the neighbors (Luce & Perry, 1949). The global clustering coefficient of the network is the average of the clustering coefficients of all nodes.

$$\text{clustering coefficient}(v) = \frac{2k}{d(d-1)} \quad (9)$$

The closeness centrality of node  $v$  measures the distance from  $v$  to every other node  $t$  in the network. It is computed as shown in Equation (10), where  $SP(v, t)$  is the length of the shortest path between  $v$  and  $t$ , and  $V$  is the set of nodes of the network (Sabidussi, 1966).

$$\text{closeness centrality}(v) = \frac{1}{\sum_{t \in V} SP(v, t)} \quad (10)$$

The betweenness centrality of node  $v$  measures the proportion of shortest paths in the network that go through  $v$ . It is computed as shown in Equation (11), where  $SP_{st}$  is the number of shortest paths between nodes  $s$  and  $t$  and  $SP_{st}(v)$  is the number of

shortest paths between  $s$  and  $t$  that pass through  $v$  (Freeman, 1977).

$$\text{betweenness centrality}(v) = \sum_{s \neq v \neq t \in V} \frac{SP_{st}(v)}{SP_{st}} \quad (11)$$

For each of the three measures, we compute the corresponding “spectrum” as the average over all nodes of degree  $k$ , for each value of  $k$ . For example, clustering spectrum of a network is the average clustering coefficient of all nodes of degree  $k$  in the network (Supplementary Figure 3). Spectra are often displayed in log scale for ease of interpretation.

### 3 Results

We constructed networks using each combination of edge weighting method, data type, and edge cut-off. We then compared networks of a given size constructed from the same data type but using different edge weighting methods. We additionally compared networks of a given size constructed using the same edge weighting method but from different data types. This was done for each of the edge cut-offs.

#### 3.1 Effects of network construction on functional content of networks

In this section, we aim to answer whether the choice of edge weighting method, data type, and edge cut-off affects the functional content of networks.

##### 3.1.1 Different networks uncover different amounts of functional content

We measured the functional content of networks in terms of known interactions and shared GO terms (see Methods). These are commonly used approaches to evaluate the extent to which networks reflect existing biological knowledge (De Smet & Marchal, 2010; Ucar et al., 2007; Nayak et al., 2009; Ashburner et al., 2000; Faith et al., 2007; Marbach et al., 2010). For example, there is a relationship between the similarity of genes’ expression patterns and interaction of the proteins that the genes encode (Grigoriev, 2001).

Specifically, we computed precision, recall, and the  $F$ -score in a network as follows. Precision is the proportion of edges in the network that correspond to known interactions or whose end nodes share a GO term. Recall is the proportion of the known interactions or gene pairs sharing a GO term that are in the network. Because there is a trade-off between precision and recall, in the sense that higher precision means lower recall and vice versa, the two measures were combined into  $F$ -score, their harmonic mean (see Methods). We computed precision, recall, and  $F$ -score for each combination of edge weighting method and data type, varying edge cut-off from 0% to 100%.

We found that the choice of *edge weighting method* affected the functional content in terms of both known interactions (when known interactions are combined independent of interaction type) and shared GO terms. Networks constructed with correlation nearly always had higher precision and recall (Figure 1), as well as  $F$ -score (Supplementary Figure 1) than networks constructed with mutual information. This indicates that correlation-based networks more accurately uncover existing

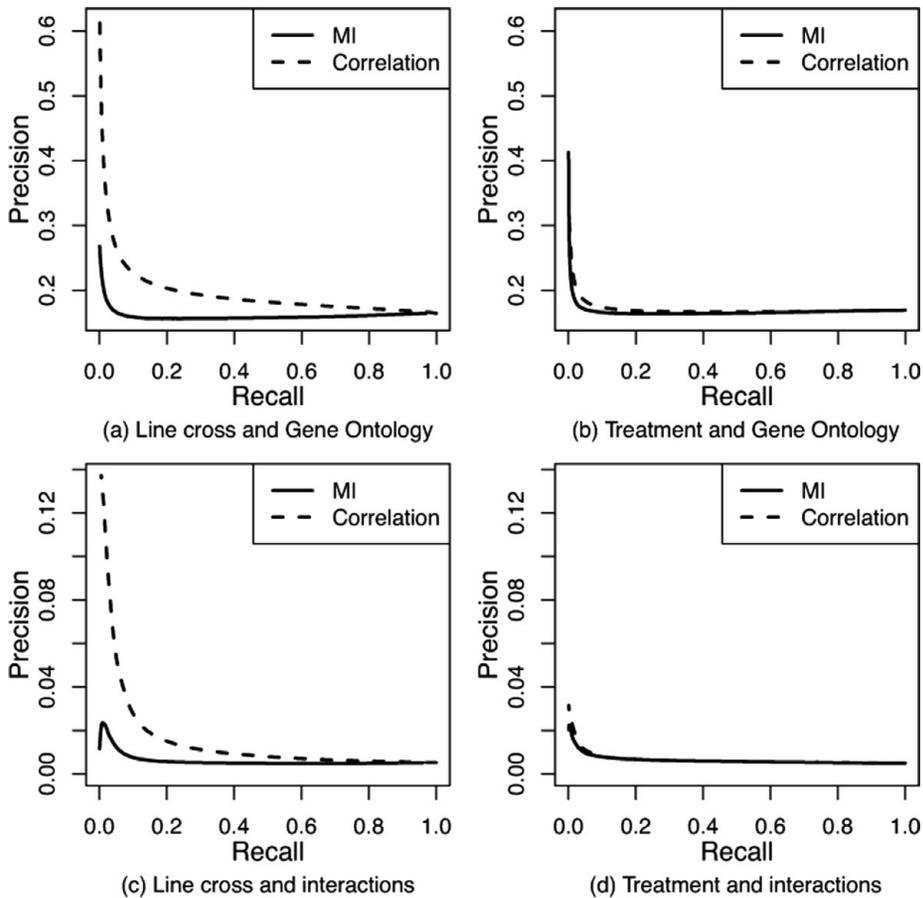


Fig. 1. Precision–recall curves measuring how accurately networks constructed in different ways capture known biological knowledge. Panel (a) shows curves for networks constructed from the line cross data with respect to shared GO terms. Panel (b) shows curves for networks constructed from the treatment data with respect to shared GO terms. Panel (c) shows curves for networks constructed from the line cross data with respect to known interactions. Panel (d) shows curves for networks constructed from the treatment data with respect to known interactions.

biological knowledge, where the level of their superiority varies depending on the edge cut-off (Figure 1 and Supplementary Figure 1). This was much more pronounced for the line cross data than for the treatment data. The choice of *data type* also affects functional content: precision and recall were higher for networks constructed from the line cross data than for networks constructed from the treatment data. This was more pronounced for correlation than for mutual information.

As expected, precision tended to decrease and recall tended to increase with an increase in edge cut-off, independent of edge weighting method or data type. This was true for both known interactions and shared GO terms. However, at smaller cut-offs, precision tended to drop significantly. Hence, the edge cut-off just before this drop in precision could be a good cut-off for constructing the network. Alternatively, one could choose the cut-off where precision and recall cross or where *F*-score starts

to decrease (e.g., see “peaks” in Supplementary Figure 1(c) and (d)). Ultimately, which edge cut-off to choose depends on one’s preference for the trade-off between precision and recall or the desired network density. We conclude that the smaller the edge cut-off, the higher the precision, the lower the recall, and the sparser the network.

Note that, whereas uncovering existing knowledge is desirable for the purpose of testing the accuracy of network construction, there is no reason to assume that gene co-expression networks should fully uncover existing knowledge, as each piece of biological data could be capturing somewhat complementary functional slices of a cell.

### 3.1.2 Different networks uncover different types of known interactions

We also evaluated known interactions in greater detail by focusing on individual interaction types, instead of considering all known interactions combined as done above. For each network and interaction type, we counted how many interactions of the given type were present in the network. We then computed the probability of observing the same or higher number of interactions of the same type in the network purely by chance using the model of hypergeometric distribution (see Methods). Here, instead of varying the edge cut-off over the entire [0%, 100%] range, as above, we varied it from 2,500 to 75,000 edges in increments of 2,500 edges, hence focusing on stronger edges only. Such a more detailed, type-specific analysis of known interactions allowed us to make more biologically relevant observations about the potential differences in the functional content between networks constructed in different ways.

We found that the choice of *edge weighting method* affected the functional content in terms of *individual* interaction types (Figures 2–4, and Supplementary Figure 2). For example, if we focus on Affinity Capture-RNA interactions, we observe that the choice of edge weighting method made a noticeable difference for the line cross data: the presence of interactions of this type was statistically significant in networks constructed using correlation at each of the 30 cut-offs, while it was not significant in networks constructed using mutual information at any of the 30 cut-offs. On the other hand, the choice of edge weighting method made a small difference for the treatment data when it came to this interaction type: the presence of interactions of this type was statistically significant in networks constructed using both correlation and mutual information at most of cut-offs, and the two edge weighting methods differed in only six out of the 30 cut-offs (Figure 4). Hence, the choice of *data type* also affected the functional content in terms of individual interaction types.

In general, the significance of the enrichment of a network constructed when using one edge weighting method is often different than when using another edge weighting method. The same holds for using one data type compared to using another data type. This observation, specifically that significance may be unique to one edge weighting approach or one data type), allowed us to systematically quantify these differences further and determine that:

1. The difference between mutual information and correlation is more pronounced in the line cross data than in the treatment data, independent of the edge cut-off. Namely, the number of known interaction types (out of 26

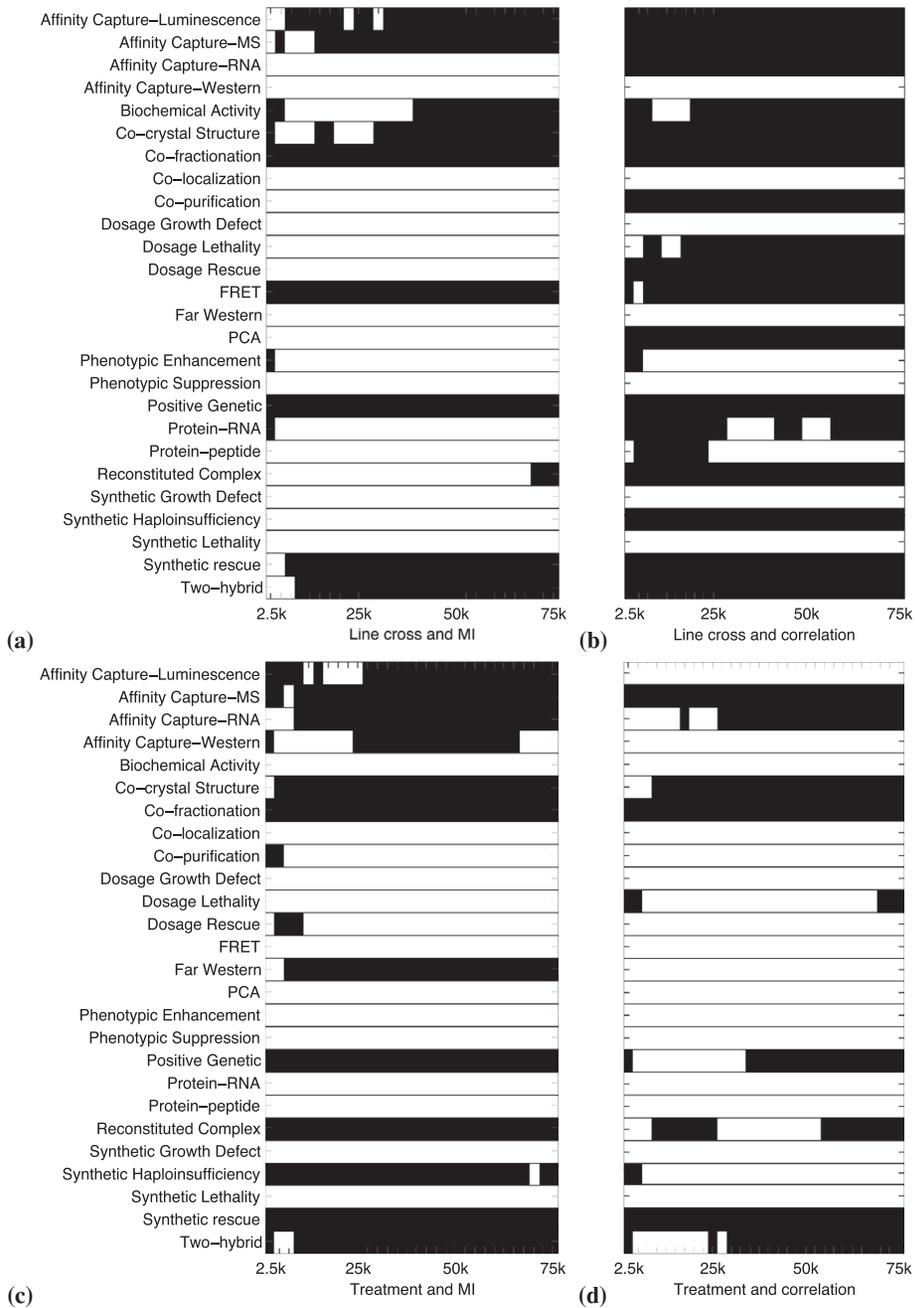


Fig. 2. Heat maps showing where the enrichment of a network of a given size (i.e., at a given edge cut-off; x-axis) in known interactions of a given type (y-axis) is statistically significant (denoted by black color) according to the hypergeometric test (see Methods). The cut-off for statistical significance was computed at a False Discovery Rate (FDR) of 0.05. Panel (a) shows enrichment results for networks constructed from the line cross data using mutual information. Panel (b) shows enrichment results for networks constructed from the line cross data using correlation. Panel (c) shows enrichment results for networks constructed from the treatment data using mutual information. Panel (d) shows enrichment results for networks constructed from the treatment data using correlation. For example, “Affinity Capture-RNA” interaction type is not significantly enriched at any edge cut-off in panel (a), it is enriched at all cut-offs in panel (b), and it is enriched at most cut-offs (particularly larger ones) in panels (c) and (d).

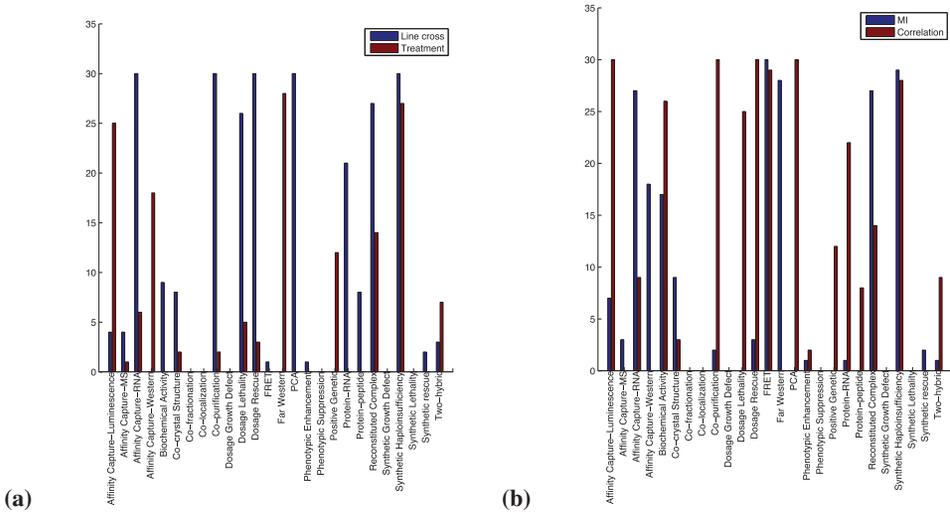


Fig. 3. The absolute differences between networks constructed using different edge weighting methods **(a)** or data types **(b)** at different edge cut-offs ( $x$ -axis) with respect to the number of known interaction types that the networks are statistically significantly enriched in ( $y$ -axis). That is, based on Figure 2, for a given data type (line cross or treatment), for a given cut-off, we count for how many of the 26 known interaction types the significance of the enrichment of the network constructed using mutual information is different from the significance of the enrichment of the network constructed using correlation (panel **(a)**). By “different” we mean that the enrichment with respect to one edge weighting method is significant (denoted by black color in Figure 2), while it is not significant with respect to the other edge weighting method (denoted by white color in Figure 2). Analogously, for a given edge weighting method (mutual information or correlation) and a given cut-off, we count for how many of the 26 known interaction types the significance of the enrichment of the network constructed from the line cross data is different than the significance of the enrichment of the network constructed from the treatment data (panel **(b)**). Now, by “different” we mean that the enrichment with respect to one data type is significant (denoted by black color in Figure 2) while it is not significant with respect to the other data type (denoted by white color in Figure 2). (color online)

of them) at which we observed a difference between the two edge weighting methods was higher for the line cross data than for the treatment data. Specifically, this number varied between 7 (27%) and 14 (54%) for line cross data and between three (12%) and eight (31%) for treatment data, depending on the edge cut-off (Figure 3(a)).

2. The difference between the line cross data and the treatment data is more pronounced for correlation than for mutual information, independent of the edge cut-off. Namely, the number of known interaction types (out of 26 of them) at which we observed a difference between the two data types was higher for correlation than for mutual information. Specifically, this number varied between 8 (31%) and 15 (58%) for correlation and between 4 (15%) and 13 (50%) for mutual information, depending on the edge cut-off (Figure 3(b)).
3. Although there are some exceptions, for each of the two results above the differences tended to decrease with an increase of the edge cut-off used (Figure 3(a) and (b)).

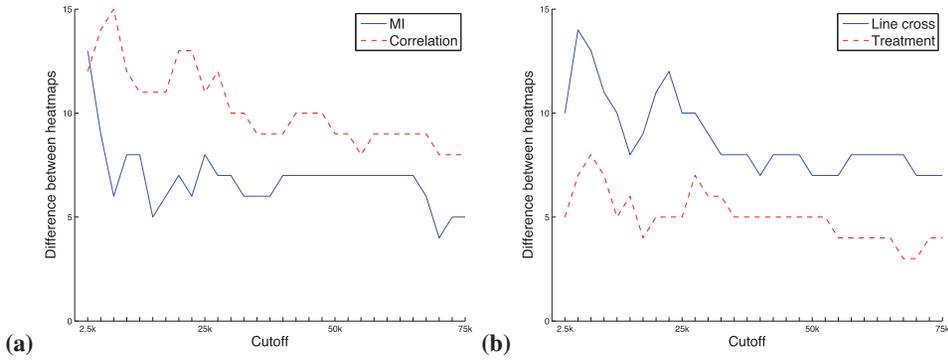


Fig. 4. The differences between networks constructed using different edge weighting methods **(a)** or data types **(b)** for different known interaction types ( $x$ -axis) with respect to the number of edge cut-offs at which the enrichment in the known interactions is different ( $y$ -axis). That is, based on Figure 2, for a given data type (line cross or treatment), for a given known interaction type, we count for how many of the 30 edge cut-offs the significance of the enrichment of the network constructed using mutual information is different than the significance of the enrichment of the network constructed using correlation (panel **(a)**). Analogously, for a given edge weighting method (mutual information or correlation) for a known interaction type, we count for how many of the 30 edge cut-offs the enrichment of the network constructed from the line cross data is different than the significance of the enrichment of the network constructed from the treatment data (panel **(b)**). (color online)

We reached the above conclusions by counting, for each cut-off, at how many of the 26 known interaction types the significance of the network constructed using one edge weighting method and one data type is different than the significance of the network constructed using either the different edge weighting method or the different data type. We reached the same conclusions when we counted, for each known interaction type, at how many of the 30 edge cut-offs we observe the difference in significance of the networks constructed in different ways, as follows:

1. The difference between mutual information and correlation is more pronounced in the line cross data than in the treatment data. Namely, mutual information and correlation were different at more cut-offs in the line cross data than in the treatment data (Figure 4(a)).
2. The difference between the line cross data and the treatment data is more pronounced for correlation than for mutual information. Namely, the line cross data and the treatment data were different at more cut-offs for correlation than for mutual information (Figure 4(b)).
3. The majority of the known interaction types show disagreement between the line cross data and the treatment data. Namely, only 11 out of the 26 known interaction types showed agreement between the line cross data and the treatment data in the sense that the number of cut-offs in which correlation and mutual information differ was similar for the two data types.
4. The majority of the known interaction types show disagreement between correlation and mutual information. Namely, only 11 out of the 26 known interaction types showed agreement between correlation and mutual information

in the sense that the number of cut-offs in which the line cross data and the treatment data differ was similar for the two edge weighting methods.

Supplementary Table 1 further supports the above conclusions: for the majority (18–21 out of 26) of known interaction types, there is a statistically significant difference ( $p\text{-value} \leq 1.7 \times 10^{-03}$ ) between the different edge weighting methods and data types with respect to the proportion of known interactions that they uncover across networks (corresponding to the 30 different cut-offs).

### 3.1.3 Different networks target different biological questions

In addition to *quantifying* the differences in the functional content between networks constructed using different edge weighting methods or data types, as above, we also *qualitatively* studied which combination of edge weighting method and data type can (or fails to) capture a given known interaction type. For each known interaction type, we checked whether its enrichment was significant in the majority of the 30 networks (corresponding to the 30 cut-offs) constructed using the given edge weighting method and data type. We then grouped known interaction types into those that were significantly enriched for both correlation and mutual information in both the line cross data and the treatment data, those that were significantly enriched for both correlation and mutual information in the line cross data but only for mutual information in the treatment data, those that were significantly enriched for both correlation and mutual information in the treatment data but only for correlation in the line cross data, ..., and those that were significantly enriched for neither correlation nor mutual information in neither the line cross data nor the treatment data (Table 1).

Only six out of the 26 known interaction types were found by both edge weighting methods in both data types. Specifically, if one's goal was to construct networks that would enrich for Affinity Capture-MS, Co-crystal Structure, Co-fractionation, Positive Genetic, Synthetic Rescue, or Two-hybrid interactions, one could use either of the two edge weighting methods or data types (Table 1). Note however, that the actual interactions uncovered by the different networks could be different, because the overlap between the networks is small (Section 3.2.1). Seven of the 26 known interaction types are missed by both edge weighting methods in both data types. In other words, neither the two edge weighting methods nor data types considered in this study enriched for Co-localization, Dosage Growth Defect, Phenotypic Enhancement, Phenotypic Suppression, Protein-peptide, Synthetic Growth Defect, or Synthetic Lethality interactions.

Clearly for  $6+7 = 13$  out of the 26 known interaction types, all four combinations of edge weighting methods and data types *agree*. For the same number of known interaction types, however, at least two of the four combinations *disagree*. For example, there exist interaction types that were captured by correlation in the line cross data but not in the treatment data. Also, there exist interaction types captured by mutual information in the treatment data but not in the line cross data. Interestingly, there are no interaction types captured by mutual information in the line cross data but not in the treatment data or by correlation in the treatment data but not in the line cross data (Table 1).

Table 1. *Combinations of edge weighting methods/data types for which the given interaction type is significantly enriched at the majority of edge cut-offs.*

Combinations of edge weighting methods/data types	Interaction types
Found by both correlation and mutual information in both the line cross data and the treatment data.	Affinity Capture-MS
	Co-crystal structure
	Co-fractionation
	Positive genetic
	Synthetic rescue
Found by both correlation and mutual information in the line cross data but only by mutual information in the treatment data.	Two-hybrid
	Affinity Capture-luminescence
Found by both correlation and mutual information in treatment data but only by correlation in line cross data.	Affinity Capture-RNA reconstituted complex
Found by both correlation and mutual information in line cross data but none of correlation or mutual information in treatment data.	Biochemical activity
	FRET
Found only by correlation in line cross data and only by mutual information in treatment data.	Synthetic haploinsufficiency
Found by neither correlation nor mutual information in treatment data, but found only by correlation in line cross data.	Co-purification
	Dosage lethality
	Dosage rescue
	PCA
	Protein-RNA
Found by neither correlation nor mutual information in line cross data, but found only by mutual information in treatment data.	Affinity Capture-western
	far western
Found by neither correlation nor mutual information in neither line cross data nor treatment data.	Co-localization
	Dosage growth defect
	Phenotypic enhancement
	Phenotypic suppression
	Protein-peptide
	Synthetic growth defect
	Synthetic lethality

### 3.1.4 Bottom line

All of the above results demonstrate that networks constructed in this study with different combinations of edge weighting methods or data types were enriched with different functional content. Therefore, we have strong evidence that both the edge weighting method and the type of biological experiment underlying the data affect the functional content of networks, implying that they can be optimized to answer certain—but different—biological questions.

## 3.2 Effects of network construction on topology of networks

In this section, we aim to answer whether the choice of edge weighting method, data type, and edge cut-off affects the topology of constructed networks.

Table 2. Edge overlap between networks constructed using different data types and edge weighting methods, averaged over all edge cut-offs. Each column denotes the edge weighting method or data type used as the basis for comparison between networks. The column denoted by “Line cross” compares networks constructed from the line cross data using correlation to networks constructed from the line cross data using mutual information. The column denoted by “Treatment” compares networks constructed from the treatment data using correlation to networks constructed from the treatment data using mutual information. The column denoted by “Correlation” compares networks constructed from the line cross data using correlation to networks constructed from the treatment data using correlation. The column denoted by “MI” compares networks constructed from the line cross data using mutual information to networks constructed from the treatment data using mutual information.

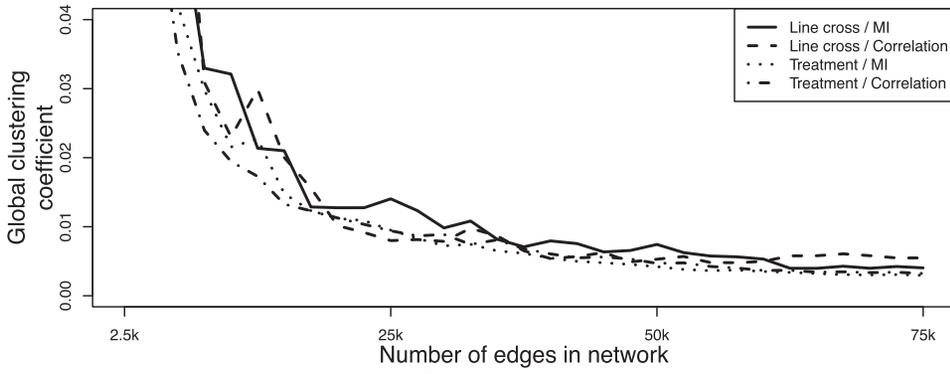
	Line cross	Treatment	Correlation	MI
Mean	30.85%	40.23%	11.98%	14.06%
Minimum	29.41%	32.08%	9.23%	11.70%
Maximum	33.12%	45.82%	13.37%	18.82%
Standard deviation	0.51%	4.30%	1.35%	2.13%

### 3.2.1 Overlap of different networks is small

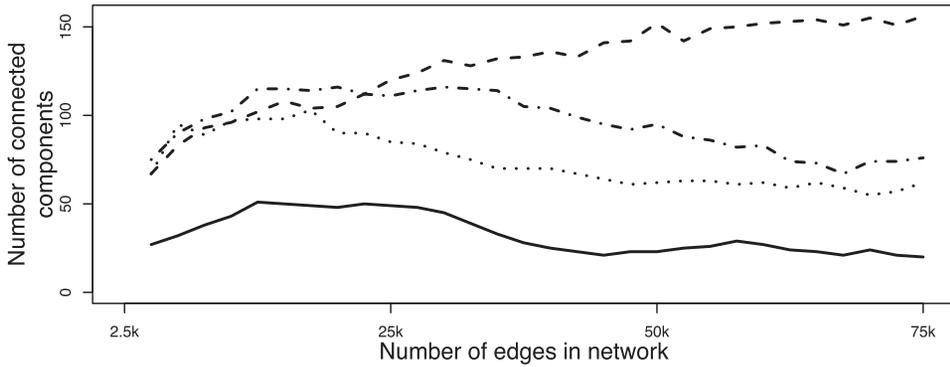
Considerable topological differences between networks constructed in different ways were immediately apparent based on the intersection of their edges (Table 2). Intersections were calculated between pairs of networks of a given size that shared either edge weighting method or data type. Because the size of the intersection between the networks appeared to have a linear relationship with the cut-off (networks with more strict cut-offs had nearly the same proportion of overlap as networks with less strict cut-offs), the intersections were averaged over different network sizes. Such averaged intersections were smaller between networks constructed by using the same edge weighting method but different data sets than between networks constructed by using the same data set but different edge weighting method (Table 2). This indicates that the choice of the data set may have a stronger effect on network construction than the choice of the edge weighting method. Nonetheless, the intersection was relatively small between all compared networks. This is an important result, since it is likely that different edges would lead to different biological interpretations, which is exactly what we demonstrated in the previous section.

### 3.2.2 Different networks have different topologies

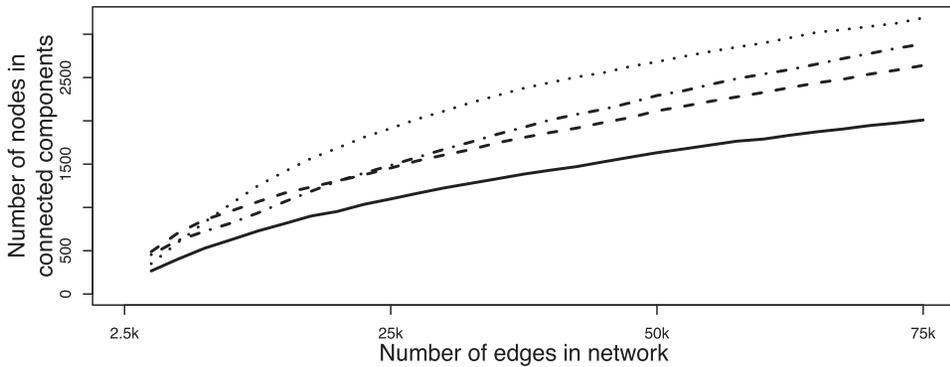
By comparing overall topological characteristics of networks constructed using different edge weighting methods or data types, we found that networks varied as follows. In general, average clustering coefficients (see Methods) were higher for networks constructed using mutual information than for networks constructed using correlation, independent of the data type, and they were higher for networks constructed from the line cross data than for networks constructed from the treatment data, independent of the edge weighting method (Figure 5(a)). Figure 5(b) and (c) further demonstrate topological differences between networks constructed with different edge weighting methods and data types. For example, networks for the line cross data and correlation had the most connected components (Figure 5(b)) but they had the second fewest nodes involved in the components (Figure 5(c)),



(a) Global clustering coefficient



(b) Number of connected components



(c) Number of nodes in connected components

Fig. 5. Global clustering coefficient **(a)**, the number of connected components with at least two nodes **(b)**, and the number of nodes that participate in the connected components **(c)** for networks constructed from a given data type (line cross or treatment) using a given edge weighting method (correlation or mutual information) at a given edge cut-off ( $x$ -axis).

indicating that these networks had many small components. On the other hand, networks for the treatment data and mutual information in general had the second fewest connected components (Figure 5(b)) but they had the most nodes involved in the components (Figure 5(c)), indicating that these networks had few large components.

To further understand the observed topological differences, we focused on one of the 30 analyzed edge cut-offs: 25,000. We chose the cut-off of 25,000 because it

is a compromise between the the peak precision in Figure 1(a)–(d). We compared the different networks corresponding to this cut-off with respect to the clustering coefficient, closeness centrality, and betweenness centrality spectra (see Methods). Topological differences were immediately apparent. For example, in the line cross data-based networks, average clustering coefficients tended to decrease as node degrees increased for mutual information, whereas this was not necessarily the case for correlation (Supplementary Figure 3(a)). Clustering spectra were also different for the line cross data-based networks (Supplementary Figure 3(a)) and treatment-based networks (Supplementary Figure 3(b)): correlation-based networks in particular appear to have almost the opposite trends for the two data types. The observed differences were statistically significant for each pair of clustering spectra, both when the data type was shared but the edge weighting method was different ( $p$ -values  $< 2.2 \times 10^{-16}$ ) and when the edge weighting method was the same but the data type was different ( $p$ -values of 0.008 and  $< 1.07 \times 10^{-15}$ ) (Supplementary Figure 3).

Similar observations were made with respect to closeness centrality spectra. For example, in the line cross data-based networks, while average closeness centralities were higher for correlation than for mutual information for low-degree nodes, they were lower for correlation than for mutual information for high-degree nodes (Supplementary Figure 4(a)). While the observed difference between the two edge weighting methods was not statistically significant for the line cross data ( $p$ -value of 0.3511), it was statistically significant for the treatment data ( $p$ -value of 0.0063; Supplementary Figure 4(b)). Further, the difference between closeness spectra was statistically significant when the edge weighting method is the same but the data type is different, with respect to both correlation ( $p$ -value of  $2.81 \times 10^{-6}$ ) and mutual information ( $p$ -value of  $1.6 \times 10^{-12}$ ) (Supplementary Figure 4). Betweenness centrality spectra are statistically significantly different for all four combinations of edge weighting methods and data types as well (Supplementary Figure 5).

We conclude that networks constructed using different edge weighting methods or data types have different topologies.

#### 4 Discussion

The network construction problem is analogous to the problem of clustering genes based on similarity between their expression profiles (Hanisch et al., 2002; Datta & Datta, 2003; Eisen et al., 1998). Many clustering algorithms exist (each with its (dis)advantages (Fortunato 2010; Solava et al. 2012; Ho et al. 2010)) that attempt to group together genes that have similar expression with respect to some distance metric. Hence, analogous to a network construction method, a clustering method first computes distances (or equivalently, similarities) between each pair of genes. Then, it partitions the resulting weighted fully connected network by employing a distance cut-off to determine cluster membership (Hartigan & Wong, 1979). Examples of popular clustering methods in this context include Walktrap (Pons & Latapy, 2005) and Markov Clustering (MCL) (Wittkop et al., 2007). The wide variety of available algorithms typically result in different clustering solutions, for reasons that are analogous to those studied in this paper.

One goal of this work was to study the effect of the choice of edge weighting method on the functional content and topology of the resulting network. We note

that this is not the first study to analyze how different edge weighting methods affect biological content of the networks. For example, it has been found that modules in networks constructed with signed correlation show different functional enrichment than modules in networks constructed with the absolute correlation (Mason et al., 2009), and that correlation and mutual information are strongly related in the treatment data when their values are high (Steuer et al., 2002). This is the first study, however, to analyze how the choice of edge weighting method affects networks in terms of *individual* interaction types. This study is also more comprehensive than the previous ones because it examines multiple edge weighting methods *and* data types.

Our hypothesis is that the two edge weighting methods might be capturing different types of relationships. For example, Supplementary Figure 6 represents a known Biochemical Activity interaction for which the gene pair ranking is discordant, i.e., high with respect to one edge weighting method but low with respect to the other edge weighting method. Supplementary Figure 7 and Supplementary Figure 8, on the other hand, represent known Synthetic Lethality interactions in the line cross data and the treatment data, respectively, for which the gene pair rankings are consistent across the two edge weighting methods—both rank each of these gene pairs high. Note that in our study we focus on answering whether different edge weighting methods captured different functional content to bring this important issue to the attention of the community, not on answering why different types of known interactions were captured differently by the different networks. The later is an important but complex question that is out of the scope of this study.

Another goal of this work was to study whether the type of biological experiment underlying the data affects the functional content of inferred networks because different types of biological experiments are expected to capture somewhat complementary aspects of the cell (Jansen, 2001; Hughes et al., 2000). To our knowledge this is the first study explicitly comparing networks constructed from line cross data with networks constructed from treatment data. We note that since we use one data set of each type, and since we do so for one species, our results should be used with caution: it is difficult to determine how much of the variation between the different networks is due to the data type versus noise in these particular data sets. Nonetheless, our analysis could be used as a *strong suggestion* that data types considered should also be carefully chosen to match the desired biological question(s).

Finally, we studied the effect of edge cut-off on resulting networks. There are a number of approaches to choosing a cut-off for network construction, including the precision–recall trade-off (Section 3.1.1), the approximate topology of the resulting network, or the false discovery rate (FDR) (Carlson et al., 2006; Ucar et al., 2007). Under the assumption that gene co-expression networks should capture known biological knowledge, be it interactions, GO functional similarity, or anything else, the strongest relationships should hold the most reliable known biological information (which would be reflected in higher precision at lower edge cut-offs).

We demonstrated systematically and comprehensively that the four combinations of edge weighting methods and data types disagreed with respect to the functional content and topology of the corresponding networks. Because the overlap was small between networks constructed in different ways, and since their overall topologies were different, it is no surprise that the different networks led to different biological

interpretations. We showed that: (1) of the two edge weighting methods, correlation seemed to more accurately uncover existing biological knowledge, especially for the line cross data; (2) of the two data types, the line cross data seemed to more accurately uncover existing biological knowledge, especially for correlation; (3) the strongest edges indeed hold the most reliable known biological information; (4) the difference between mutual information and correlation was more pronounced in the line cross data than in treatment data, independent of the edge cut-off; (5) the difference between the line cross data and the treatment data was more pronounced for correlation than for mutual information, independent of the edge cut-off; (6) the differences in points (4) and (5) above tended to decrease with an increase of the cut-off; (7) the type of experiment underlying the data can have at least as much of an effect on the functional content of networks as the choice of edge weighting method; and (8) different types of known interactions could be uncovered with different combinations of edge weighting methods and data types.

In summary, these results demonstrate that there is no single correct way to construct a network—and different approaches can produce different networks that are better suited to answer different biological questions. For example, using correlation and treatment data results in networks that contain the most significant amounts of Synthetic Lethality interactions among all considered networks. Thus, this combination of edge weighting method and data type may be the most appropriate for studying genes that are essential to an organism's survival in combination with other genes. However, the same networks are the worst in terms of capturing Far Western interactions, which are a type of PPI dataset. As such they should probably not be used to study molecular processes that are carried out via physical interactions between proteins. We conclude that the three ubiquitous factors in network inference that were considered in this study each have significant effects on the functional content and topology of resulting biological networks and therefore do *matter* for future work in systems biology.

### Acknowledgments

Research was supported in part by the National Science Foundation (NSF) grants BCS-0826958 and EAGER CCF-1243295.

### Supplementary materials

For supplementary material for this article, please visit <http://dx.doi.org/10.1017/nws.2014.13>

### References

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, **25**(1), 25–29.
- Barabasi, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, **5**(2), 101–113.
- Brem, R. B., & Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(5), 1572–1577.

- Butte, A. J., & Kohane, I. S. (2000). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In R. Altman, K. Dunker, L. Hunter, K. Lauderdale, & T. Klein (Eds.), *Pacific symposium for biocomputing*, Vol. 5 (pp. 418–429). Hawaii.
- Carlson, M., Zhang, B., Fang, Z., Mischel, P., Horvath, S., & Nelson, S. (2006). Gene connectivity, function, and sequence conservation: Predictions from modular yeast co-expression networks. *BMC Genomics*, *7*(1), 40.
- Christie, K. R., Hong, E. L., & Cherry, J. M. (2009). Functional annotations for the *Saccharomyces cerevisiae* genome: The knowns and the known unknowns. *Trends in Microbiology*, *17*(7), 286–294.
- Datta, S., & Datta, S. (2003). Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, *19*(4), 459–466.
- Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*, 1–30.
- De Smet, R., & Marchal, K. (2010). Advantages and limitations of current network inference methods. *Nature Reviews Microbiology*, *8*(10), 717–729.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 14863–14868.
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., . . . Gardner, T. S. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, *5*(1), e8.
- Feizi, S., Marbach, D., Medard, M., & Kellis, M. (2013). Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature Biotechnology*, *31*(8), 726–733.
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174.
- Freeman, L. C. (1977). A set of measures of centrality. *Sociometry*, *40*(1), 35–41.
- Grigoriev, A. (2001). A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *29*(17), 3513–3519.
- Hanisch, D., Zien, A., Zimmer, R., & Lengauer, T. (2002). Co-clustering of biological networks and gene expression data. *Bioinformatics*, *18*(Suppl. 1), S145–S154.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108.
- Ho, H., Milenković, T., Memišević, V., Aruri, J., Pržulj, N., & Ganesan, A. (2010). Protein interaction network uncovers melanogenesis regulatory network components within functional genomics datasets. *BMC Systems Biology*, *4*(84).
- Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., . . . Friend, S. H. (2000). Functional discovery via a compendium of expression profiles. *Cell*, *102*(1), 109–126.
- Ivliev, A. E., AC't Hoen, P., & Sergeeva, M. G. (2010). Coexpression network analysis identifies transcriptional modules related to proastrocytic differentiation and sprouty signaling in glioma. *Cancer Research*, *70*(24), 10060–10070.
- Jansen, R. (2001). Genetical genomics: The added value from segregation. *Trends in Genetics*, *17*(7), 388–391.
- Kok, S., & Domingos, P. (2005). Learning the structure of Markov logic networks. In De Raedt & S. Wrobel (Eds.), *Proceedings of the 22nd International Conference on Machine Learning* (pp. 441–448). Bonn, Germany: ACM.
- Landgrebe, T. C. W., Paclik, P., Duin, R. P. W., & Bradley, A. P. (2006). Precision-recall operating characteristic (P-ROC) curves in imprecise environments. In Y. Y. Tang, S. P. Wang, G. Lorette, D. S. Yeung, & H. Yan (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition*, Vol. 4, pp. 123–127. Hong Kong.
- Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, *14*(2), 95–116.

- Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D., & Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(14), 6286–6291.
- Margolin, A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R., & Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, **7**(Suppl. 1), S7.
- Markowetz, F., & Spang, R. (2007). Inferring cellular networks – A review. *BMC Bioinformatics*, **8**(Suppl. 6).
- Mason, M., Fan, G., Plath, K., Zhou, Q., & Horvath, S. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*, **10**(1), 327.
- Meyer, P., Lafitte, F., & Bontempi, G. (2008). MINET: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, **9**(1), 461.
- Milenković, T., Lai, J., & Przulj, N. (2008). GraphCrunch: A tool for large network analyses. *BMC Bioinformatics*, **9**(1), 70.
- Milenković, T., Memišević, V., Bonato, A., & Pržulj, N. (2011). Dominating biological networks. *PLOS ONE*, **6**(8), e23016.
- Nayak, R. R., Kearns, M., Spielman, R. S., & Cheung, V. G. (2009). Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Research*, **19**(11), 1953–1962.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **33**(3), 1065–1076.
- Pe'er, D., Regev, A., Elidan, G., & Friedman, N. (2001). Inferring subnetworks from perturbed expression profiles. *Bioinformatics*, **17**(Suppl. 1), S215–S224.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(8), 1226–1238.
- Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. In plnar Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and Information Sciences – ISCIS 2005*, Vol. 3733, Chap. 31 (pp. 284–293). Berlin/Heidelberg: Springer.
- Rider, A. K., Siwo, G., Emrich, S. J., Ferdig, M. T., & Chawla, N. V. (2014). A supervised learning approach to the ensemble clustering of genes. *International Journal of Data Mining and Bioinformatics*, **9**(2), 199–219.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, **31**(4), 581–603.
- Smith, E. N., & Kruglyak, L. (2008). Gene-environment interaction in yeast gene expression. *PLoS Biology*, **6**(4), e83.
- Solava, R. W., Michaels, R. P., & Milenković, T. (2012). Graphlet-based edge clustering reveals pathogen-interacting proteins. *Bioinformatics*, **18**(28), i480–i486. Also, in Proceedings of the 11th European Conference on Computational Biology (ECCB), Basel, Switzerland, September 9–12, 2012 (acceptance rate: 14%).
- Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, **18**(Suppl. 2), S231–S240.
- Ucar, D., Neuhaus, I., Ross-MacDonald, P., Tilford, C., Parthasarathy, S., Siemers, N., & Ji, R. R. (2007). Construction of a reference gene association network from multiple profiling data: application to data analysis. *Bioinformatics*, **23**(20), 2716–2724.
- van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., . . . Bernard, R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, **347**(25), 1999–2009.
- van Noort, V., Snel, B., & Huynen, M. A. (2004). The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Reports*, **5**(3), 280–284.
- van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., . . . Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**(6871), 530–536.

- Wittkop, T., Baumbach, J., Lobo, F. P., & Rahmann, S. (2007). Large scale clustering of protein sequences with force-a layout based heuristic for weighted cluster editing. *BMC Bioinformatics*, **8**(1), 396.
- Zhou, X., Kao, M.-C. C., & Hung, W. (2002). Transitive functional annotation by shortest-path analysis of gene expression data. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(20), 12783–12788.
- Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., Bumgarner, R. E., & Schadt, E. E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, **40**(7), 854–861.