J. R. Soc. Lond. Interface (0000)
DOI: 10.1098/rsif0000.0000

Modelling Protein-Protein Interaction Networks via a Stickiness Index

Nataša Pržulj[1†] and Desmond J. Higham[2,]

[1]*Department of Computer Science, University of California, Irvine, CA, 92697-3425, USA*
[2]*Department of Mathematics, University of Strathclyde, Glasgow, G11XH, Scotland, UK*

What type of connectivity structure are we seeing in protein-protein interaction networks? A number of random graph models have been mooted. After fitting model parameters to real data, the models can be judged by their success in reproducing key network properties. Here, we propose a very simple random graph model that inserts a connection according to the degree, or "stickiness", of the two proteins involved. This model can be regarded as a testable distillation of more sophisticated versions that attempt to account for the presence of interaction surfaces, or binding domains. Computing a range of network similarity measures, including relative graphlet frequency, we find that our model outperforms other random graph classes. In particular, we show that given the underlying degree information, fitting a stickiness model produces better results than simply choosing a degree-matching graph uniformly at random. The results therefore lend support to the basic modelling methodology.

Keywords: protein-protein interaction networks; network models

1. INTRODUCTION AND MODEL

A protein-protein interaction (PPI) network is commonly viewed as an unweighted, undirected graph. Each node in the graph represents a protein and an edge between a pair of nodes indicates that those proteins have been observed to interact physically [23, 43, 18, 26, 37, 41]. The types of connectivity patterns that arise are neither completely random, in the classical Erdös-Rényi sense, nor completely deterministic [20].

In an attempt to understand and describe the PPI connectivities, a number of models, that is, formulas for generating edges in some probabilistic sense, have been proposed and tested against observed networks [24, 27, 3, 36, 9]. Much work has focussed on matching degree distributions and recovering a scale-free law [24, 27, 3, 38] although whether PPI networks are really scale-free is still the subject of debate [12, 16, 25, 21, 36]. Our aim here is to present a new, pared-down, but biologically motivated model that simplifies previous work to the extent that fitting parameters and comparing local and global graph properties becomes meaningful and revealing.

Among the few existing models that incorporate some biological justification are those of Caldarelli et al. [8], Thomas et al. [42] and Deeds et al. [10]. These related models have in common the idea that proteins interact because they share complimentary physical aspects, a concept that is consistent with the underlying biochemistry. Following [42] we will refer to these physical aspects as binding domains. The approach in those papers is to generate graphs by assigning binding domain information to the nodes at random and then inserting links probabilistically according to some pairwise matching criterion. The aim is then to reproduce properties observed in real PPI networks, most notably the degree distribution. We also mention that a refined "lock-and-key" version of the model from [42] has been used to extract protein-level detail from real data sets [30], further justifying the modelling approach.

Currently, it would be a very challenging task to infer from a real PPI network the number and distribution of distinct binding domains [4, 11], not least because the networks are known to be noisy [40]. For this reason, it is difficult to decide whether the models from [8, 10, 42] are being tested under realistic parameter ranges. We therefore propose a simplified model that attempts to summarize the abundance and popularity of binding domains on a protein as a single number based on its normalized degree; we call this number the *stickiness index*. The model has the benefit of being tunable to the given degree structure of a PPI network. In this way, a benchmark model that captures the essence of [8, 10, 42] can be tested.

Our work can be motivated by two main assumptions.

Assumption 1. Having a high degree implies that a protein has many binding domains and/or its binding domains are commonly involved in interactions.

Assumption 2. A pair of proteins is more likely to interact (share complementary binding domains) if they both have high stickiness indices, and correspondingly less likely to interact if one or both have a low stickiness index. We thus take the product of the two stickiness indices to define the probability of interaction—this borrows from the concept of

†Author for correspondence (natasha@ics.uci.edu).

an AND gate in Boolean logic [5] and the idea of a rank-one approximation in dimension reduction [13].

The following pseudocode defines our model.

```
input  {deg_i}_{i=1}^N, list of degrees of N nodes
output {w_ij}_{i,j=1}^N, adjacency matrix from model

for i = 1 to N
    θ_i = deg_i / sqrt(Σ_{j=1}^N deg_j)
end
Initialize all w_ij = 0
for i = 1 to N
    for j = i to N
        compute a uniform (0,1) sample, r
        if r ≤ θ_i θ_j
            w_ij = 1 and w_ji = 1
        end if
    end for
end for
```

This choice of stickiness index θ_i ensures that the ith node in the model has *expected* degree $\deg_i$. Moreover, under Assumption 2, this definition of stickiness in terms of degree is the only one that captures the correct expected degree. Details are given in the appendix.

Our stickiness index coincides with the concept of *fitness* in [8] with the notable distinction that fitness in [8] is assigned at random, with a focus on the resulting degree distribution, whereas stickiness above is assigned deterministically, based on the unique choice that matches expected degrees. Since we do not require any other parameter fitting, this approach allows us to perform a 'proof of principle' test of the basic idea that links can be modelled via mutual compatibility.

Note that high degree proteins in current PPI networks may not necessarily contain a plentiful amount of binding domains, as implied by our Assumption 1. Instead, their high connectivities may be artifacts of *technical false positives*, auto-activators or "sticky" proteins, or due to *biological false positives*, as some PPIs can occur in the experimental procedure, but not *in vivo* because protein pairs are not expressed at the same time, in the same sub-cellular compartment, or in the same tissue [21]. Thus, our Assumption 1 may be a severe over-simplification for some proteins in current PPI data sets. Nevertheless, as PPI detection biotechnologies improve to produce cleaner, higher-confidence PPI data, Assumption 1 will become more descriptive of the observed networks.

A multitude of random graph models that reproduce scale-free degree distributions have been proposed, although the relevance of scale-freeness to PPI networks has been questioned [12, 16, 25, 21, 36]. The most notable such models are those based on biologically motivated *gene duplication and mutation* network growth principles [19, 46, 33, 44]. In these models, networks grow by duplication of nodes (genes), and as a node gets duplicated, it inherits most of the neighbors (interactions) of the parent node, but gains some new neighbors as well. Thus, a hybrid model having properties of both a gene duplication–mutation model and the stickiness index based model is a promising future direction. In such a model, a duplicated gene would inherit the parent's stickiness index along with many of the parent's neighbors, as in a gene duplication–mutation model and it would gain new neighbors in proportion to its inherited stickiness index and stickiness indices of the nodes already in the network, as in our stickiness index based model.

We remark that early tests on low confidence data in [27] suggest that PPI networks have a bias against connections between high degree proteins. This is potentially at odds with the models in [8, 10, 42], where sets of proteins that share matching and commonly occuring (high fitness) physical aspects will interact and all have high degree. In our simple model we assign edges independently, but it would be possible to add a post-processing stage in which links were rewired in order to test for various types of correlation. Hence, a further application of our model is in studying correlation effects in PPI network topology.

2. EXPERIMENTS AND RESULTS

Comparing large real-world networks is computationally intensive as it involves an NP-complete *subgraph isomorphism problem* [47]. Thus, simple heuristics measuring *global* and *local* network properties, have been used. The most commonly examined global network properties are the *degree distribution*, *clustering coefficient*, and network *diameter* (see [32] for a detailed survey). More recently, bottom-up local approaches to studying a network structure have been proposed [29, 39, 36]. Analogous to sequence motifs, *network motifs* have been defined as subgraphs that recur in a network at frequencies much higher than those found in randomized networks [29, 39, 28]; they were used to uncover basic functional units in various real-world networks. To account for frequencies of occurence of all small subgraphs rather than for only the over-represented ones, *graphlets* were defined as small connected non-isomorphic induced subgraphs of a large network and their *relative frequencies* were used to define a new *distance* measure between two networks [36].

To examine the fit of our new stickiness index based model of PPI networks, we use all of these standard global and local network parameters. The relative graphlet frequency distance is the most demanding network similarity measure, imposing 29 different constraints on the networks being compared (details in [36]), so we use it as our main comparison tool. We compared fourteen large publicly available PPI networks with sample networks from five models, including the stickiness model.

We used PPI networks of the following eukaryotic organisms: yeast *S. cerevisiae*, fruitfly *D. melanogaster*, nematode worm *C. elegans*, and human. Several different data sets are available for yeast and human, so we analyzed five yeast PPI networks of different confidence levels obtained from three different high-throughput studies [43, 23, 45], as well as five human PPI networks obtained from the two recent high-throughput studies [41, 37] and three curated data bases [1, 35, 48]. We denote by "YHC" the high-confidence yeast PPI

network from [45], by "Y11K" the yeast PPI network defined by the top 11, 000 interactions in the von Mering *et al.* classification [45], by "YIC" the Ito *et al.* "core" yeast PPI network [23], by "YU" the Uetz *et al.* yeast PPI network [43], and by "YICU" the union of "YIC" and "YU" yeast PPI networks (we unioned them as in [21] to increase coverage). "FE" and "FH" denote the fruitfly *D. melanogaster* entire and high-confidence PPI networks from [18]. Similarly, "WE" and "WC" denote the worm *C. elegans* entire and "core" PPI networks from [26]. Finally, "HS", "HR", "HB", "HH", and "HM" stand for human PPI networks from yeast two-hybrid (Y2H) screens by Stelzl *et al.* [41] and Rual *et al.* [37], and from curated databases BIND [1], HPRD [35], and MINT [48], respectively (BIND, HPRD, and MINT data were downloaded from OPHID [7] on February 10, 2006). Note that YHC and Y11K networks are mainly coming from tandem affinity purifications (TAP) [17] and high throughput mass spectrometry protein complex identification (HMS-PCI) [22], while YIC, YU, YICU, FE, FH, WE, WH, HS, and HR are yeast two-hybrid, and HB, HH, and HM are a result of human curation (BIND, HPRD, and MINT). Thus, we are using PPI networks of different confidence levels that come from a range of high throughput PPI detection biotechnologies as well as from human curation.

We compared these PPI networks with the following five model networks: Erdös-Rényi random graphs [14, 15] (henceforth denoted by "ER"), random graphs with exactly the same degree distribution as a PPI network [6, 31] (denoted "R-SF" for "random scale-free"), Barabasi-Albert scale-free networks [2] (denoted by "BA-SF"), 3-dimensional geometric random graphs [34] (denoted by "GEO-3D"), and the stickiness model networks described above (denoted by "STICKY").

For each of the fourteen PPI networks, and for each of the five models, we compared the PPI network with 25 samples from the model. Each sample matched the number of nodes and edges in the corresponding PPI network.

Average relative graphlet frequency distances between the PPI and the corresponding model networks for each of the five network models are presented in Figure 1. The stickiness model shows an improved fit over all other network models with respect to relative graphlet frequency distances in ten out of the fourteen tested PPI networks (black squares in Figure 1); it fits as well as the GEO-3D model (white squares in Figure 1) in one and is outperformed by the GEO-3D model in three PPI networks. In addition, this model reproduces global network properties such as the degree distribution (see the appendix), the clustering coefficients (white circles in Figure 2, left), and the average diameters of PPI networks (white circles in Figure 2, right).

It is of particular note that the R-SF model does not perform as well as the stickiness model. This means that, given the degree distribution of a PPI network,

(a) simply drawing a network uniformly at random from the class of all networks that match the degree distribution is less successful at capturing the underlying substructure than

(b) enhancing this degree information by using the simple modelling insights summarized in Assumptions 1 and 2.

3. CONCLUSIONS

Overall, the stickiness framework produces a convenient, parameter-free random network that is motivated by transparent modelling arguments and may be regarded as a simplified, testable, distillation of more sophisticated models. The results give further justification for the modelling approaches in [8, 10, 42]. Since the model accurately reproduces all widely used quantitative measures it also provides a benchmark against which others may be compared.

Acknowledgement We thank the referees for valuable feedback.

4. APPENDIX

Suppose $A \in \mathbb{R}^{N\times N}$ is the PPI network adjacency matrix, so $a_{ij} = a_{ji} = 1$ if proteins i and j are connected and $a_{ij} = a_{ji} = 0$ otherwise. We are using $\deg_i := \sum_{j=1}^{N} a_{ij}$ to denote the degree of protein i.

Suppose that some function of the degree, $f^{[i]}(\deg_i)$, defines the stickiness index of protein i. Then, under Assumption 2 (and independently for each distinct pair of proteins)

$$\mathbb{P}(i \leftrightarrow j) = f^{[i]}(\deg_i) \cdot f^{[j]}(\deg_j),$$

where $i \leftrightarrow j$ denotes the event that i and j are connected.

In order to match the PPI network degree with the expected degree from the model, we require

$$\begin{aligned} \deg_i &= \mathbb{E}[\text{degree of node } i \text{ in model}] \\ &= \sum_{j=1}^{N} \mathbb{P}(i \leftrightarrow j) \\ &= \sum_{j=1}^{N} f^{[i]}(\deg_i) \cdot f^{[j]}(\deg_j) \\ &= f^{[i]}(\deg_i) \sum_{j=1}^{N} f^{[j]}(\deg_j). \end{aligned}$$

Let $C = \sum_{j=1}^{N} f^{[j]}(\deg_j)$. Then the formula above tells us that $\deg_i = C f^{[i]}(\deg_i)$, and thus

$$f^{[i]}(\deg_i) = \frac{\deg_i}{C}.$$

Summing over i shows that $C^2 = \sum_{i=1}^{N} \deg_i$. We conclude that

$$f^{[i]}(\deg_i) = \frac{\deg_i}{\sqrt{\sum_{j=1}^{N} \deg_j}},$$

confirming that our stickiness index θ_i is uniquely defined under our assumptions.

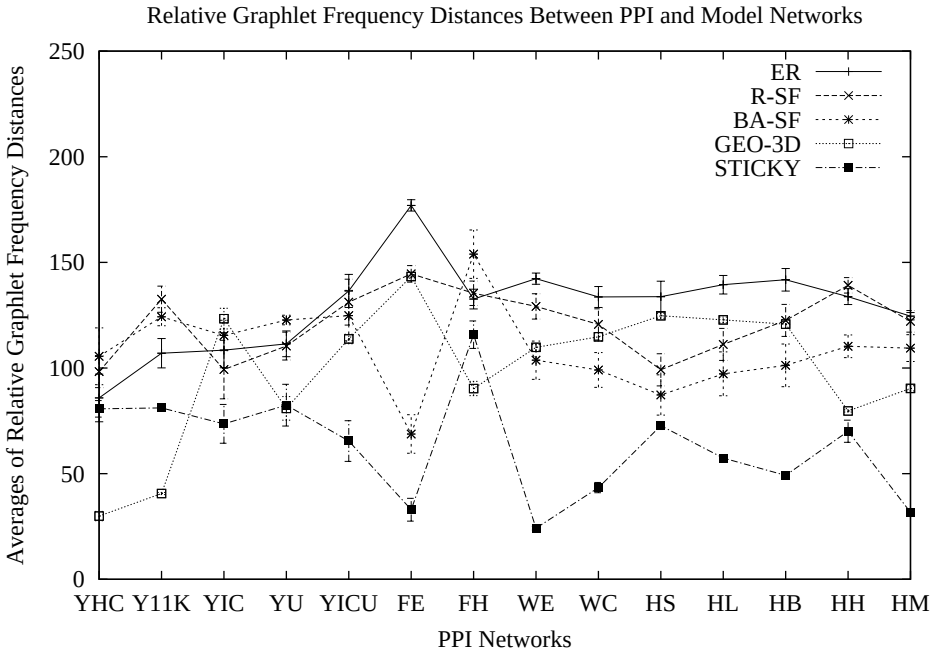

Figure 1. Relative graphlet frequency distances (y-axis) between the fourteen PPI networks (x-axis) and their corresponding model networks. The lower the number, the better the fit. Averages of distances between 25 sample networks and the corresponding PPI network are presented for each random graph model and each PPI network. Points are joined for clarity only. The error bar around a point spans one standard deviation above and below (in some cases, error bars are barely visible, since they are of the size of the point). Labels on the horizontal axis are described in the text.

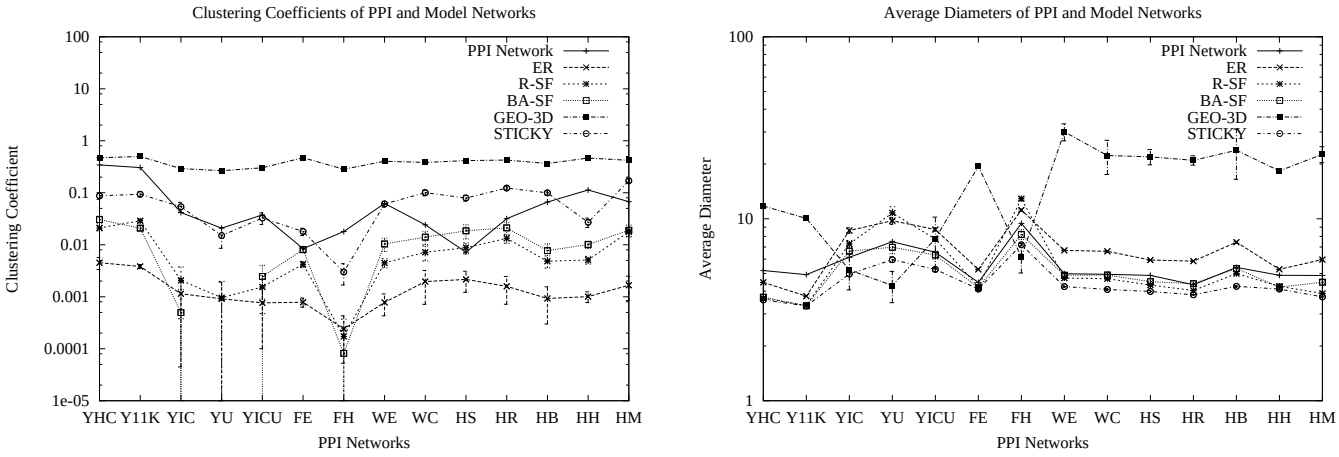

Figure 2. **Left:** Clustering coefficients of fourteen PPI networks and averages of clustering coefficients of 25 model networks corresponding to a PPI network. **Right:** Average diameters of the fourteen PPI networks and averages of average diameters of 25 model networks corresponding to a PPI network. Error bars and labels are described in the legend of Figure 1.

We note that in order for all probabilities to be in the range $[0, 1]$, we require $\theta_i\theta_j \leq 1$ for all i, j. (Assuming that all proteins have at least one interaction, a sufficient condition is that the product of the two largest degrees is bounded by N). This property held for all networks considered here.

As discussed in [8], an intuitively reasonable alternative to the multiplicative model is the additive version

$$\mathbb{P}(i \leftrightarrow j) = g^{[i]}(\deg_i) + g^{[j]}(\deg_j).$$

However, copying the same style of analysis leads to the conclusion that

$$g^{[i]}(\deg_i) = \frac{\deg_i}{N} - \frac{1}{2N}\sum_{k=1}^{N} \deg_k,$$

so that

$$\mathbb{P}(i \leftrightarrow j) = \frac{1}{N}\left(\deg_i + \deg_j - \frac{1}{N}\sum_{k=1}^{N} \deg_k\right).$$

Since many proteins have degree less than half the network average, this model breaks down due to the assignment of negative probabilities.

REFERENCES

[1] G. D. Bader, D. Betel, and C. W. V. Hogue. BIND: the biomolecular interaction network database. *Nucleic Acids Research*, 31(1):248–250, 2003.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–12, 1999.

[3] A.-L. Barabási, Z. Dezso, E. Ravasz, Z.-H. Yook, and Z. N. Oltvai. Scale-free and hierarchical structures in complex networks. In *Modeling of Ccomplex Systems: Seventh Granada Lectures. AIP Conference Proceedings*, volume 661, pages 1–16, 2003.

[4] A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and K. L. L. Sonnhammer. The pfam protein families database. *Nucleic Acids Research*, 30(1):276–280, 2002.

[5] M. Ben-Ari. *Mathematical Logic for Computer Science*. Springer, 2001.

[6] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory A*, 24:296–307, 1978.

[7] K. Brown and I. Jurisica. Online predicted human interaction database. *Bioinformatics*, 2005.

[8] G. Caldarelli, A. Capocci, P. De Los Rios, and M.A. Mu noz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89:258702–1–4, 2002.

[9] E. de Silva and M.P.H. Stumpf. Complex networks and simple models in biology. *Roy. Soc. Interface*, 2:419–430, 2005.

[10] Eric J. Deeds, Orr Ashenberg, and Eugene I. Shakhnovich. A simple physical model for scaling in protein-protein interaction networks. *Proceedings of the National Academy of Sciences*, 103:311–316, 2006.

[11] M. Deng, S. Mehta, F. Sun, and T. Chen. Inferring domain-domain interactions from protein-protein interactions. *Genome Research*, 12(10):1540–1548, 2002.

[12] D. Dupuy, N. Bertin, M. E. Cusick, J.-D. J. Han, and M. Vidal. Reply to toward the complete interactome. *Nature Biotechnology*, 24(6):615–615, 2006.

[13] Lars Eldén. *Matrix Methods in Data Mining and Pattern Recognition*. SIAM, PA, 2006 (to appear).

[14] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[15] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.

[16] C. C. Friedel and R. Zimmer. Toward the complete interactome. *Nature Biotechnology*, 24(6):614–615, 2006.

[17] A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–7, 2002.

[18] L Giot, JS Bader, C Brouwer, A Chaudhuri, B Kuang, Y Li, YL Hao, CE Ooi, B Godwin, E Vitols, G Vijayadamodar, P Pochart, H Machineni, M Welsh, Y Kong, B Zerhusen, R Malcolm, Z Varrone, A Collis, M Minto, S. Burgess, L McDaniel, E Stimpson, F Spriggs, J Williams, K. Neurath, N Ioime, M Agee, E Voss, K Furtak, R Renzulli, N Aanensen, S Carrolla, E Bickelhaupt, Y Lazovatsky, A DaSilva, J Zhong, CA Stanyon, RL Jr Finley, KP White, M Braverman, T Jarvie, S Gold, M Leach, J Knight, RA Shimkets, MP McKenna, J Chant, and JM Rothberg. A protein interaction map of drosophila melanogaster. *Science*, 302(5651):1727–1736, 2003.

[19] K.-I. Goh, B. Kahng, and D. Kim. Hybrid network model: the protein and the protein family interaction networks. *arXiv:q-bio.MN/0312009 v2, 28 March 2004*, 2004.

[20] P. Grindrod and M. Kibble. Review of uses of network and graph theory concepts within proteomics. *Expert Rev. Proteomics*, 1(2):89–98, 2004.

[21] J. D. H. Han, D. Dupuy, N. Bertin, M. E. Cusick, and Vidal. M. Effect of sampling on topology predictions of protein-protein interaction networks. *Nature Biotechnology*, 23:839–844, 2005.

[22] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Rasmussen, J. R. Andersen, L. E. Johansen, L. H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthiesen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys, and M. Tyers. Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry. *Nature*, 415(6868):180–3, 2002.

[23] T. Ito, K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci U S A*, 97(3):1143–7, 2000.

[24] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–2, 2001.

[25] Raya Khanin and Ernst Wit. How scale-free are gene networks? *Journal of Computational Biology*, 13(3):810–818, 2006.

[26] S Li, CM Armstrong, N Bertin, H Ge, S Milstein, M Boxem, P-O Vidalain, J-DJ Han, A Chesneau, T Hao, N Goldberg, DS Li, M Martinez, J-F Rual, P Lamesch, L Xu, M Tewari, SL Wong, LV Zhang, GF Berriz, L Jacotot, P Vaglio, J Reboul, T Hirozane-Kishikawa, Q Li, HW Gabel, A Elewa, B Baumgartner, DJ Rose, H Yu, S Bosak, R Sequerra, A Fraser, SE Mango, WM Saxton, S Strome, S van den Heuvel, F Piano, J Vandenhaute, C Sardet, M Gerstein, L Doucette-Stamm, KC Gunsalus, JW Harper, ME Cusick, FP Roth, DE Hill, and M Vidal. A map of the interactome network of the metazoan c. elegans. *Science*, 303:540–543, 2004.

[27] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–3, 2002.

[28] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.

[29] R. Milo, S. S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298:824–827, 2002.

[30] J. L. Morrison, R. Breitling, D. J. Higham, and D. R. Gilbert. A lock-and-key model for protein-protein interactions. *Bioinformatics*, 2006. Advance Access published online on June 20, 2006; doi:10.1093/bioinformatics/btl338.

[31] M. E. J. Newman. Random graphs as models of networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*. Wiley-VHC, Berlin, 2002.

[32] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[33] R. Pastor-Satorras, E. Smith, and R. V. Sole. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, 222:199–210, 2003.

[34] M. Penrose. *Geometric Random Graphs*. Oxford Univeristy Press, 2003.

[35] S. Peri, J. D. Navarro, T. Z. Kristiansen, R. Amanchy, V. Surendranath, B. Muthusamy, T. K. Gandhi, K. N. Chandrika, N. Deshpande, S. Suresh, B. P. Rashmi, K. Shanker, N. Padma, V. N iranjan, H. C. Harsha, N. Talreja, B. M. Vrushabendra, M. A. Ramya, A. J. Yatish, M. Joy, H. N. S hivashankar, M. P. Kavitha, M. Menezes, D. R. Choudhury, N. Ghosh, R. Saravana, S. Chandran, S. Mohan, C. K. Jonnalagadda, C. K. Prasad, C. Kumar-Sinha, K. S. Deshpande, and A. Pandey. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res*, 32 Database issue:D497–501, 2004. 1362-4962 Journal Article.

[36] N. Pržulj, D. G. Corneil, and I. Jurisica. Modeling interactome: Scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.

[37] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437:1173–78, 2005.

[38] Marcel Salathé, Robert M. May, and Sebastian Bonhoeffer. The evolution of network topology by selective removal. *Roy.Soc.Interface*, 2:533–536, 2005.

[39] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31:64–68, 2002.

[40] E. Sprinzak, S. Sattath, and H. Margalit. How reliable are experimental protein-protein interaction data? *Journal of Molecular Biology*, 327:–919–923, 2003.

[41] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E.E. Wanker. A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, 122:957–968, 2005.

[42] A. Thomas, R. Cannings, N. A. M. Monk, and C. Cannings. On the structure of protein-protein interaction networks. *Biochemical Society Transactions*, 31:1491–1496, 2003.

[43] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, E. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleish, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. *Nature*, 403:623–627, 2000.

[44] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Modeling of protein interaction networks. *ComPlexUs*, 1:38–44, 2001.

[45] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887):399–403, 2002.

[46] A. Wagner. How the global structure of protein interaction networks evolves. *Proceedings of The Royal Society of London. Series B, Biological Sciences*, 270:457–466, 2003.

[47] D. B. West. *Introduction to Graph Theory.* Prentice Hall, Upper Saddle River, NJ., 2nd edition, 2001.

[48] A. Zanzoni, L. Montecchi-Palazzi, M. Quondam, G. Ausiello, Helmer-Citterich M., and G. Cesareni. Mint: A molecular interaction database. *FEBS Letters*, 513(1):135–140, 2002.