How threshold behaviour affects the use of subgraphs for network comparison

Tiago Rito[1,2,*], Zi Wang[1], Charlotte M. Deane[1,2] and Gesine Reinert[1,2]

[1]Department of Statistics, University of Oxford, OX1 3TG and [2]OCISB/DTC Systems Biology, Department of Biochemistry, South Parks Road, OX1 3QU, UK

ABSTRACT

Motivation: A wealth of protein–protein interaction (PPI) data has recently become available. These data are organized as PPI networks and an efficient and biologically meaningful method to compare such PPI networks is needed. As a first step, we would like to compare observed networks to established network models, under the aspect of small subgraph counts, as these are conjectured to relate to functional modules in the PPI network. We employ the software tool GraphCrunch with the Graphlet Degree Distribution Agreement (GDDA) score to examine the use of such counts for network comparison.

Results: Our results show that the GDDA score has a pronounced dependency on the number of edges and vertices of the networks being considered. This should be taken into account when testing the fit of models. We provide a method for assessing the statistical significance of the fit between random graph models and biological networks based on non-parametric tests. Using this method we examine the fit of Erdös–Rényi (ER), ER with fixed degree distribution and geometric (3D) models to PPI networks. Under these rigorous tests none of these models fit to the PPI networks. The GDDA score is not stable in the region of graph density relevant to current PPI networks. We hypothesize that this score instability is due to the networks under consideration having a graph density in the threshold region for the appearance of small subgraphs. This is true for both geometric (3D) and ER random graph models. Such threshold behaviour may be linked to the robustness and efficiency properties of the PPI networks.

Contact: tiago@stats.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Recent advances in experimental science and in literature mining techniques have generated a considerable amount of protein–protein interaction (PPI) data from several organisms. These interactions are often integrated to form networks, which can help put the proteins into their functional and physiological context. A reliable and efficient method for large network comparison would be very useful (Sharan and Ideker, 2006). Such a comparison may yield mechanistic and evolutionary insights, help to identify missing links and even aid network validation for those organisms where experimental data are scarce. A first step is to establish a method for network comparison with well-studied random network models.

Current PPI networks are unfortunately still very incomplete and rife with noise (von Mering *et al.*, 2002). They tend to have a

large number of false positives and false negatives. These obscure meaningful conclusions and offer challenges to robust methods of analysis (Alm and Arkin, 2003). Network comparison is also a computationally difficult task because typical PPI networks are relatively large, e.g. *Saccharomyces cerevisiae* already has about 18 440 binary PPIs in the DIP™ database.

Depending on the aspect under which networks are to be compared, short lists of summary statistics are often used. Classical summary statistics used include the degree distribution, the mean path length and the clustering coefficient (see Costa *et al.*, 2007 for an overview and many more summary statistics). Here, we compare networks based on small subgraphs, since cell biology is thought of as modular; many pathways and feedback loops are inherently seen as detachable modules (Hartwell *et al.*, 1999). While it has been shown that network motifs alone do not determine function in general (Ingram *et al.*, 2006), there is the possibility of a close connection between subgraphs and biological functionality (Shen-Orr *et al.*, 2002).

Our aim is to compare biological networks and random graph models under the aspect of similar subgraph counts. Such subgraph counts were introduced by Milo *et al.* (2002) with the aim of detecting over-represented small subgraphs. They compared counts for connected 3–4 node subgraphs in real-world networks to those of certain random networks, and called those patterns *network motifs*; see also Ciriello and Guerra (2008) for a review.

Counting small connected subgraphs in large PPI networks is computationally demanding. Moreover, the number of possible subgraphs of *n*-nodes increases exponentially with *n*, e.g. for $n=3$ we have two differently connected subgraphs, and 21 for $n=5$. Przulj *et al.* (2004) disregarded the frequency subjacent to the definition of motifs and counted connected induced subgraphs with 3–5 nodes, which they call *graphlets* (Fig. 1). A subgraph S of G is said to be *induced* if it contains all the edges that appear in G over the same subset of nodes. For example, the only induced subgraphs of a triangle are edges. Methods have also been developed to count 6- and 7-node graphlets (Hormozdiari *et al.*, 2007, Grochow and Kellis, 2007). Alon *et al.* (2008) used a combinatorial colour-coding technique to count up to 10-node non-induced subgraphs, arguing that these are more relevant to compare the incomplete and noisy networks currently available.

To combine the distributions that result from these graphlet counts, the so-called *relative graphlet frequency (RGF) distance* (Przulj *et al.*, 2004) and the *Graphlet Degree Distribution Agreement (GDDA)* have been suggested (Przulj, 2007). The RGF distance identifies all subgraphs with 3–5 nodes in two networks and compares the frequency of their appearance, while the GDDA statistic defines node-specific permutation groups, called automorphism orbits, within each of the 29 (2–5 nodes) possible

---

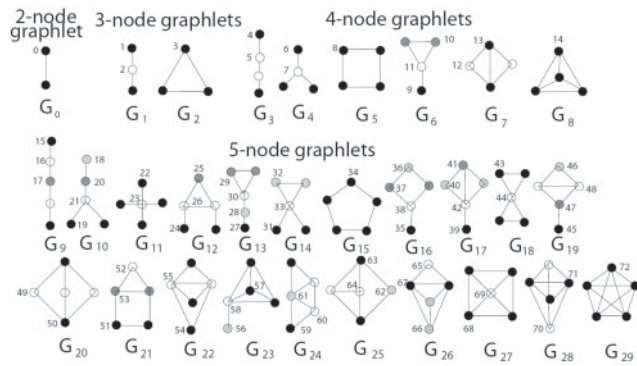*To whom correspondence should be addressed.

Fig. 1. Graphlets with 2–5 nodes $G_0, G_1, ..., G_{29}$. The automorphism orbits are numbered 0 to 72 and the nodes belonging to the same orbit are of the same shade within a graphlet. From Przulj (2007).

graphlets of the two networks being compared. Their scaled and normalized *orbit degree distributions* are then reduced by averaging the Euclidean distances between matching orbits of each network over all orbits (see Section 2 for a detailed description).

In this article, we use the software tool GraphCrunch (Milenkovic *et al.*, 2008) to examine the use of graphlet counts for network comparison. Focusing our analysis on the more refined GDDA, we find the statistic to have a non-monotone dependency on the number of edges and nodes of the networks being considered. As suggested in Przulj (2007), we use the GDDA score to compare PPI networks with three random graph models: Erdös–Rényi (ER) random graphs, ER random graphs with fixed degree distribution and geometric three-dimensional (GEO3D) random graphs. Observing that the empirical distribution of the GDDA score under these theoretical models is far from normal, we provide non-parametric test procedures to assess the model fit. We find that none of the random network models considered fit the PPI networks. While we conclude that we are still far from having a satisfactory null model for PPI networks, we provide a statistical framework for assessing the fit of potential new models under the aspect of similarity of small subgraph counts. The proposed method relies only on the assumption that if a PPI network is generated by a given model, then the empirical distributions of the GDDA comparisons between the PPI network versus model, and between model versus model, will be similar. Hence, any future model proposed for PPI networks can also be tested using this method.

Strikingly, the GDDA score is not stable in the graph density region of the biological networks considered. We hypothesize that this instability arises because the observed graph densities fall in the threshold regions for the appearance of small subgraphs, under both ER and GEO3D random graph models. In this region there is high volatility in subgraph counts even for two networks which are generated under the same model and with the same specifications. While neither of these models fit the data, we can still use their threshold regions as proxy and conjecture that the PPI networks under consideration operate near the threshold for the appearance of small subgraphs. Such behaviour would imply relatively short paths between proteins in networks, with presumably just enough alternative paths to ensure robustness, while maintaining a low edge density. This behaviour may also have further implications in the optimal design of networks.

2 MATERIALS AND METHODS

2.1 PPI networks considered

PPI networks are modelled by an undirected graph whose nodes represent the individual proteins; an edge is drawn between every two proteins which are known to interact. Multiple edges and self-loops are excluded. Six PPI networks were analysed: three of yeast, *Saccharomyces cerevisiae* and three of human, *Homo sapiens* (Table 1).

BioGRID interaction data for human (release 2.0.55, www.thebiogrid.org) was filtered using the key words 'Affinity Capture-MS' and 'Two-hybrid' and divided into two distinct datasets: BG_MS and BG_Y2H, respectively.

2.2 Basic graph theory definitions

An undirected graph G with no loops or multiple edges is a pair $(V(G), E(G))$ where the elements of $V(G)$ represent the set of vertices; the elements of $E(G)$ are called edges, and they are two-element subsets $\{v, w\}$ of $V(G)$. When $\{v, w\} \in E(G)$ we say v and w are adjacent. The *degree* of a vertex v, $\deg(v)$, is the number of edges which have v as one of its endpoints. If $V(G)$ has v elements and $E(G)$ has e elements, then the *average degree* of a graph is defined as $d(G) = 2e/v$.

A subgraph of $G = (V, E)$ is a graph $F = (V', E')$ whose vertex set $V' \subseteq V$ and its edge set $E' \subseteq E$ connects only nodes of V'. The *maximum average degree*, $m(G)$, of a graph G is the largest average degree over all subgraphs of G. A subgraph F of G is said to be *induced* by V' if and only if it includes all the edges of G which connect the vertices of V', i.e. for each pair of vertices in F and their corresponding pair in G, there will be an edge between a pair of vertices in F if there is an edge between the corresponding pair in G. Two graphs are said to be *isomorphic* if there is a one-to-one mapping f between the vertex sets of G and H such that vertices v and w are adjacent in H if and only if $f(v)$ and $f(w)$ are adjacent in G. For more background on random graphs, see for example Bollobás (2001).

In this article, we define the *graph density* ρ of a graph G, with v vertices and e edges, as the ratio between the number of edges e and the number of potential edges of G, i.e. $\rho = e / \binom{n}{2}$.

2.3 Random graph models

In this article, we focus on the following random graph models: ER (Erdös and Rényi, 1960), ER with the same degree distribution (ER-DD) as the input graph, and GEO3D (see for example Penrose, 2003).

The ER random graph model, $G_{n,m}$, has n labelled nodes connected by m edges which are randomly chosen from the $n(n-1)/2$ possible edges (Erdös and Rényi, 1960). In this model the choice of an edge is not entirely independent of the choice of another edge (Bollobás, 2001).

ER-DD is a variation of this model, it has not just the same number of nodes and edges as the input PPI network, but also the same degree distribution.

GEO3D random graphs are constructed by assigning each node random coordinates in a 3D box of unit volume, i.e. coordinates are drawn from a uniform distribution on the unit interval (see for example Penrose, 2003). Points in the box will then correspond to graph nodes, and two nodes will be connected by an edge if the Euclidean distance between them is at most r.

2.4 Thresholds for subgraph appearances

Many theoretical properties of graphs change dramatically in a narrow range of m, which lead to the concept of *threshold functions* (Erdös and Rényi, 1960). If Q is a graph property, $P(Q)$ denotes the probability that $G(n, m)$ has or belongs to Q. We say that *almost every graph* in $G(n, M(n))$ has the property Q if $P(Q) \to 1$ as $n \to \infty$. For a given monotone increasing property Q (such as the appearance of a certain subgraph), we define a threshold function $t(n)$ for Q as any function which satisfies

$$P(Q) \to 0 \quad \text{if} \quad \frac{M(n)}{t(n)} \to 0, \qquad \text{and} \qquad P(Q) \to 1 \quad \text{if} \quad \frac{M(n)}{t(n)} \to \infty.$$

Table 1. PPI networks analysed in this study

Name	#nodes	#edges	Graph Density	Average Degree	Experiment type	Organism	Reference
YIC[a]	796	841	0.00266	2.11	Yeast two-hybrid	*Saccharomyces cerevisiae*	Ito *et al.* (2000)
YHC[a]	988	2455	0.00503	4.97	TAP MS	*Saccharomyces cerevisiae*	von Mering *et al.* (2002)
HS[a]	1705	3186	0.00219	3.74	Yeast two-hybrid	*Homo sapiens*	Stelzl *et al.* (2005)
HG[a]	3134	6725	0.00137	4.29	Yeast two-hybrid	*Homo sapiens*	Rual *et al.* (2005)
BG_MS	1923	3866	0.00209	4.02	Affinity Capture-MS	*Homo sapiens*	BIOGRID (filtered)
BG_Y2H	5057	9442	0.00074	3.73	Yeast two-hybrid	*Homo sapiens*	BIOGRID (filtered)

[a]These datasets were also considered by Przulj (2007).

Table 2. Graph density values for expecting approximately one copy of the graphlets $G_1, \ldots, G_{29}$ (Fig. 1) in ER networks with 500, 1000 and 2000 vertices

Graphlets	500	1000	2000
G_1	0.00022	0.00008	0.00003
G_2	0.00363	0.00182	0.00091
G_3-G_4	0.00073	0.00029	0.00011
G_5-G_6	0.00443	0.00221	0.00111
G_7	0.01309	0.00752	0.00432
G_8	0.02696	0.01698	0.01070
G_9-G_{11}	0.00140	0.00059	0.00025
$G_{12}-G_{16}$	0.00521	0.00261	0.00130
$G_{17}-G_{21}$	0.01251	0.00702	0.00394
$G_{22}-G_{25}$	0.02340	0.01426	0.00869
$G_{26}-G_{27}$	0.03741	0.02426	0.01573
G_{28}	0.05390	0.03667	0.02495
G_{29}	0.07218	0.05104	0.03609

The expectation values for ER graphs with 500 and 2000 vertices are represented as arrows along the *x*-axis in Figure 3A.

Table 3. Approximate graph density threshold values for the appearance of *k*-vertices graphlets in GEO3D networks with 500, 1000 and 2000 vertices

Graphlets	500	1000	2000
3-vertices	0.0030	0.0011	0.0004
4-vertices	0.0085	0.0033	0.0013
5-vertices	0.0142	0.0060	0.0025

The threshold values for the appearance of 3-vertices graphlets in GEO3D graphs with 500, 1000 and 2000 vertices are represented as arrows along the *x*-axis in Figure 3B.

the gamma function $\Gamma(x)$ (Dall and Christensen, 2002),

$$r = \frac{1}{\sqrt{\pi}} \left[\frac{\alpha}{n} \Gamma\left(\frac{d+2}{2}\right) \right]^{1/d}. \tag{1}$$

Solving for α in (1) gives the threshold graph density ρ using

$$\rho = \frac{\alpha n/2}{\binom{n}{2}}.$$

Table 3 gives threshold functions of 3-,4- and 5-vertices induced graphlets for GEO3D graphs with 500, 1000 and 2000 vertices.

2.5 GraphCrunch

The random graphs used in our experiments were generated using the internal generators of GraphCrunch. GraphCrunch (Milenkovic *et al.*, 2008) is an open source software tool that compares large real-world networks with random graph models. These are automatically generated to have the same number of nodes and edges (to within 1%) as those of the real-world network being compared. This has to be taken as approximate; with a simple 12-star as input, ER-DD graphs with 10, 11 and 12 edges are generated. As well as many global standard properties, the software supports the local statistics RGF distance and GDDA. Recently, the software has been used for a wide range of applications among which are assessing parametric models for PPI networks (Przulj, 2007), protein structure networks (Milenkovic *et al.*, 2009) and brain functional networks (Kuchaiev *et al.*, 2009).

2.6 GDDA

GDDA (Przulj, 2007) is based on *orbit degree distributions*, which are based on the automorphism orbits of the 29 graphlets on 2–5 vertices, as follows. Automorphisms are edge-preserving bijections from a graph to itself, and together they form a permutation group. An *automorphism orbit* is a node that represents this group. Within the 29 graphlets, 73 different orbits can be found (Fig. 1) and each one will have an associated orbit degree distribution. An orbit *i* from graphlet G_j has *orbit degree k* in the graph G if there are k copies of G_j in G which involve orbit *i*. In Przulj (2007) the term *graphlet degree distribution* is used instead of *orbit degree distribution*, but as orbits are counted, in our view the latter term is more appropriate. For example, considering a simple 2-star graph as our main

Threshold functions are not unique although they are so within certain factors (Bollobás, 2001, p. 40). For the random graph model $G(n, M(n))$, it is possible to show that the threshold function for the property of containing a fixed, non-empty graph F is $n^{2-2/m}$, where $m = m(F)$ is the maximum average degree of F (see Bollobás, 2001, p. 89). We relate $M(n)$ and the graph density ρ via $\rho = M(n)/\binom{n}{2}$.

For the ER model it is possible and more informative to calculate the graph density such that the expected number of copies of a given subgraph F is approximately 1. For a subgraph on v vertices with e edges, the approximate expected count for the subgraph under the ER model is

$$E(\text{number of occurrences}) = \lambda \sim n^v \rho^e / v!,$$

for small ρ. When the number of occurrences is well approximated by a Poisson random variable, as in the case for balanced graphs, $P(\text{no occurrence of subgraph}) \sim 1 - e^{-\lambda} \sim \lambda$ and hence the threshold function and the expectation formula coincide. The graph density values where the expected number of counts of a specific graphlet of Figure 1 is approximately 1 (i.e. $\lambda = 1$), are given in Table 2. The values decrease with increasing number of vertices.

Threshold functions for GEO3D models are not so well understood. One can, nonetheless, calculate approximate threshold values for the appearance of induced graphlets with k vertices. Penrose (2003) showed that for a random geometric graph placed in $\mathbb{R}^d$ with n vertices and a radius r, the k-vertices subgraph count satisfies a Poisson limit when the product $n^k r^{d(k-1)}$ tends to a finite constant. The radius r can be related to the average degree α by using
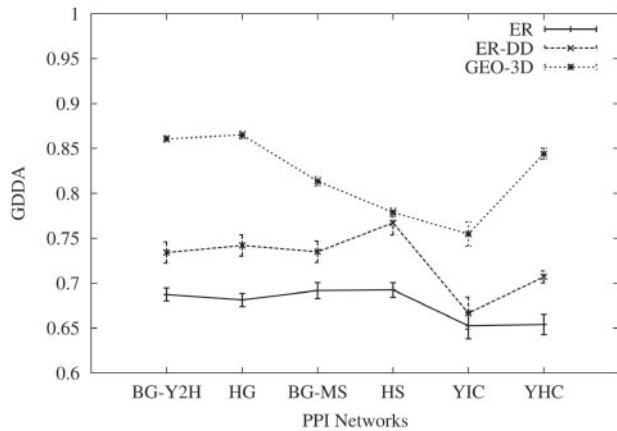
Fig. 2. GDDA plot between six PPI networks and their corresponding random model networks. All the points in the plot are averages of comparisons between the query network and each of 100 generated model networks. The error bars represent one estimated SD below and above the average point.

graph G (graphlet G_1 in Fig. 1), we would have an orbit degree distribution for orbit 0 (an edge) of two node counts for orbit degree 1 (the outer two nodes) and one count for an orbit degree 2 (the middle node); the orbit degree distribution of orbit 1 would be two counts for an orbit degree 1, and for orbit 2 we would have one count for an orbit degree 1. Let $d_G^j(k)$ be the sample distribution of the node counts for a given orbit degree k in a graph G and for a particular automorphism orbit j. In our example, where $G = G_1$, we obtain $d_{G_1}^0 = (2, 1, 0, ..., 0)$; $d_{G_1}^1 = (2, 0, 0, ..., 0)$; $d_{G_1}^2 = (1, 0, 0, ..., 0)$; and $d_{G_1}^i = 0$, for $i = 3, ..., 72$. This sample distribution is then scaled by $1/k$ in order that large degrees do not dominate the score, and normalized to give a total sum of 1,

$$N_G^j(k) = \frac{d_G^j(k)/k}{\sum_{\ell=1}^{\infty} d_G^j(\ell)/\ell}.$$

The comparison $D^j(G, H)$ of two graphs G and H with respect to j is simply the Euclidean distance between the two scaled and normalized vectors N, which is scaled by $1/\sqrt{2}$ to be between 0 and 1, as pointed out in Przulj (2010); the resulting expression is

$$D^j(G, H) = \frac{1}{\sqrt{2}} \left(\sum_{k=1}^{\infty} [N_G^j(k) - N_H^j(k)]^2 \right)^{\frac{1}{2}}.$$

This is then turned into an agreement by subtracting from 1, and the agreements are combined into a single value by taking the arithmetic mean over all j, yielding the GDDA,

$$\text{GDDA} = \frac{1}{73} \sum_{j=0}^{72} (1 - D^j(G, H)).$$

The software also calculates a variant of GDDA using the geometric mean (Supplementary Material).

A typical output based on GDDA generated by GraphCrunch is shown in Figure 2. Six PPI networks were considered; two yeast and four human. The query networks were compared with 100 random graphs of each model—ER, ER-DD and GEO—which were automatically generated by GraphCrunch.

2.7 Empirical distributions of GDDA

To address how to interpret the output from a graph comparison based on GDDA, first for both the ER model and the GEO3D model, graphs of 500, 1000 and 2000 vertices with increasing graph density were generated using the internal generators from GraphCrunch. The graphs were subsequently used as query networks in the software and compared with 50 networks of the same model, to ascertain typical GDDA scores if the model is correct.

2.8 Assessing model fit

As GDDA scores are not normally distributed in the graph density region of interest, to assess whether a given query network fits a particular model network we resort to non-parametric procedures. Given an input graph with n vertices and e edges, and a random graph model 1,

(1) Generate M graphs, say $M = 99$, from model 1 with about n vertices and e edges.

(2) For each one of these, carry out comparisons with N graphs generated from the same model and record GDDA; call the result *Sample A*. Here we use $N = 99$.

(3) Calculate the GDDA between the input graph and the N graphs from model 1, call the result *Sample B*.

(4) A histogram of *Sample A versus Sample B* may already show a clear separation of the two samples, making it obvious that the suggested model 1 is not a good fit, see Figure 4 for an illustration.

(5) For a statistical test, which tests for the null hypothesis that the two samples come from the same distributions against the general alternative that the distributions of the two samples are not the same, we employ a Monte Carlo test (see Supplementary Material for details). Here, *Sample A* records M averages of the N comparisons, whereas *Sample B* consists of one observation: the average GDDA over the N comparisons of the input network versus model. The lowest obtainable P-value is then $1/(M+1)$.

(6) We also employ a Wilcoxon rank-sum test, which tests for the alternative that the distribution of *Sample A* is a shifted version of the distribution of *Sample B* (Supplementary Material). This test is more powerful than the Monte Carlo test, but tests against a less general alternative.
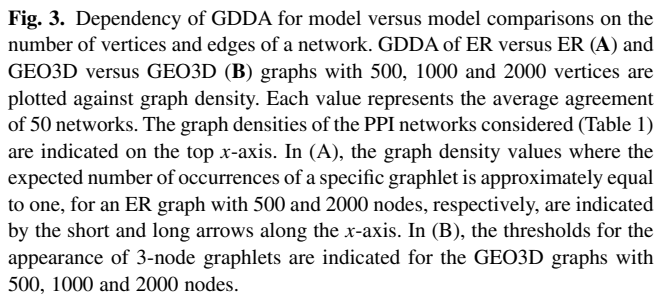
3 RESULTS AND DISCUSSION

First, PPI networks were compared with random model networks using GDDA and the standard GraphCrunch output (Fig. 2). The plot shows the highest GDDA for the GEO3D random graph model type for all the networks, followed by ER-DD and ER models. While Przulj (2007) would now conclude that GEO3D is the best fitting model for PPI networks, we shall see that due to the threshold behaviour of the networks such conclusion is not statistically justifiable.

According to Przulj (2007), a perfect score can be achieved when comparing networks of the same random model type. Przulj (2007) found the mean GDDA of comparing ER versus ER, ER-DD versus ER-DD or GEO-3D versus GEO-3D to be 0.84 ± 0.07, where 0.07 denotes one standard error. This was updated in Przulj (2010) where they found the highest score for two GEO-3D networks to be 0.95 ± 0.002.

The results for comparing ER versus ER and GEO3D versus GEO3D networks with 500, 1000 and 2000 nodes across a wide range of graph densities are summarized in Figure 3 using GDDA. Similar results for GDDA with geometric mean and for RGF-distance can be found in the Supplementary Material.

In contrast to Przulj (2007), we find that the GDDA values have not only striking differences amongst different model types but also a pronounced dependency on the number of vertices of the network. For a specific graph, drawn from one model type and with a fixed number of vertices, we also observe a strong dependency of the GDDA score with graph density when comparing to graphs of the same type and with the same number of vertices. Furthermore, these dependencies are not monotone. For easier readability and because

Fig. 3. Dependency of GDDA for model versus model comparisons on the number of vertices and edges of a network. GDDA of ER versus ER (**A**) and GEO3D versus GEO3D (**B**) graphs with 500, 1000 and 2000 vertices are plotted against graph density. Each value represents the average agreement of 50 networks. The graph densities of the PPI networks considered (Table 1) are indicated on the top *x*-axis. In (A), the graph density values where the expected number of occurrences of a specific graphlet is approximately equal to one, for an ER graph with 500 and 2000 nodes, respectively, are indicated by the short and long arrows along the *x*-axis. In (B), the thresholds for the appearance of 3-node graphlets are indicated for the GEO3D graphs with 500, 1000 and 2000 nodes.

a normal approximation does not hold, we omit the error bars from the plots of Figure 3.

In Figure 3A, for ER versus ER comparisons, in the region of graph density 0–0.01 we observe high volatility in the GDDA score, after which it increases with the graph density (Supplementary Material). This volatility may be related to the natural appearance of small subgraphs, which is itself dependent on the number of nodes. Threshold functions for the property of containing one specific graphlet were defined and calculated for ER networks with 500, 1000 and 2000 vertices (Table 2). The threshold values of the different 3–5 node graphlets for an ER graph with 500 and 2000 vertices are indicated in Figure 3A. For all graphs tested, the instability region in the GDDA score includes most of these thresholds.

For GEO3D versus GEO3D comparisons, one sees an instability in the score for small graph density which, after recovery, seems to slowly decrease again. Comparisons of GEO3D with 500 vertices for higher graph densities (up to 0.4) suggest that the score becomes more stable, although slowly increasing (Supplementary Material). While in ER graphs edges are near-independent, this is not true for GEO3D graphs because, in a geometric setting, if an edge *i* is close to *j* and if *j* is close to *k*, then *i* is likely to be close to *k* (Penrose, 2003). The asymptotic results also appear to be related to the score instability (Table 3; Fig. 3B). The most dramatic change in the score occurs when 3-node subgraphs start to appear; the appearance of 4- and 5-node subgraphs seems to have a much lower influence on the score. Strikingly, all the PPI networks under consideration are in the region of graph density populated by thresholds in both ER and GEO3D models. This invites the conjecture that PPI networks operate near the threshold for appearance of small subgraphs. Unfortunately, no good model yet exists of PPI networks and so further work will be needed to confirm this conjecture.

It is worth noting that the specific GDDA values presented in Figure 3 may vary, precisely because the specific graphs being generated for a particular comparison can be very diverse, especially in the region of high volatility (graph density between 0–0.01).

The instability of GDDA scores makes it difficult to interpret the output presented in Figure 2, not just because the typical score is different for each model type, but also because it is a function of the number of vertices and edges of the specific network being analysed. We find that the empirical distribution of GDDA in the region of interest, even in model versus model comparison, is not close to normal, indeed not even unimodal. This again can be explained by the network parameters being close to thresholds for the appearance of small subgraphs. Thus, this threshold behaviour seriously affects the statistical inference from subgraph counts for network comparison and the conclusions which can be drawn from such subgraph count comparison.

Hence, for assessing the model fit based on GDDA, we propose a new protocol. Several same model versus model comparisons with roughly the same number of vertices and edges should be carried out in order to assess the best obtainable score for this specific case. GDDA should then be calculated between the query network and graphs from the model network. Model fit can be evaluated by gauging the differences between the distributions of agreement scores resulting from query network versus model and model versus model comparisons. We suggest the Monte Carlo non-parametric test for assessing whether the two independent samples of GDDA scores, one resulting from comparisons between query network versus model and the other from model versus model comparisons, come from the same distribution. Alternatively, the Wilcoxon rank-sum test can be employed (Supplementary Material).

Figure 4A and B shows histograms of GDDA values for comparisons between the PPI network BG-MS versus 99 GEO3D and 99 ER-DD model networks, respectively. Both models have a zero Wilcoxon *P*-value (there is no overlap between the distributions). A Monte Carlo test was performed with 999 values, each an average of 30 model versus model agreements ($M=999$, $N=30$). In both cases a *P*-value of 0.001 was obtained, which is the smallest possible *P*-value for this test with 1000 observations. Although the mean of the empirical distribution is closer to ER-DD than to GEO-3D, the means are too far away to draw any useful conclusions. The large distances instead point to both
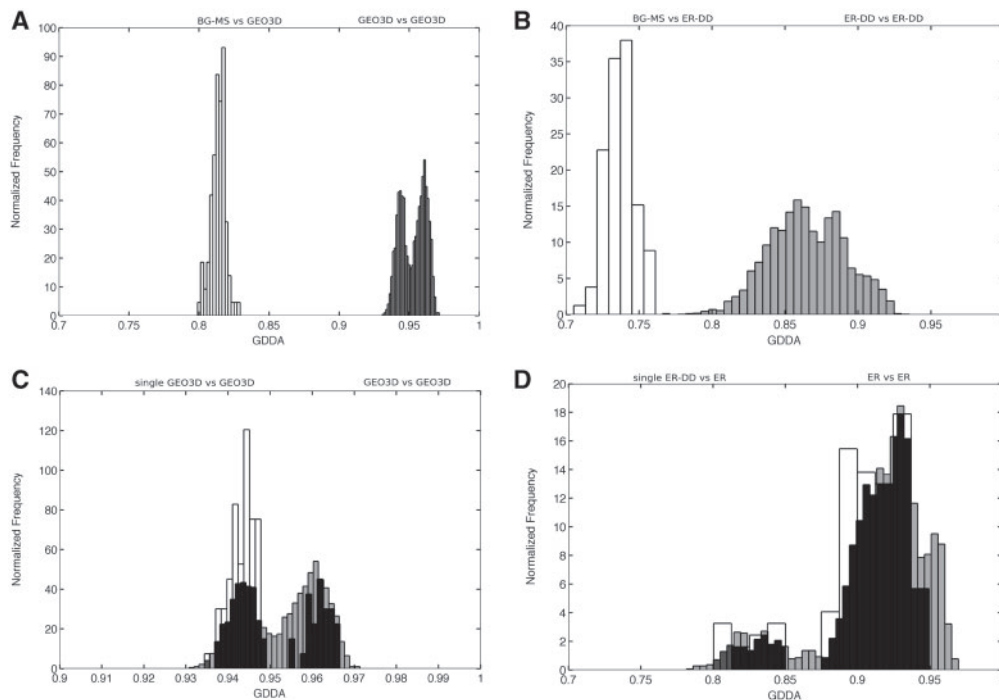
Fig. 4. Normalized histograms of GDDA values. (**A**) Histograms of GDDA values between PPI network BG-MS versus 99 GEO3D model networks *(white)* and GDDA of 30 GEO3D, each versus 99 GEO3D *(grey)* (**B**) Histograms of GDDA values between PPI network BG-MS versus 99 ER-DD model networks *(white)* and GDDA of 30 ER-DD, each versus 99 ER-DD *(grey)* (**C**) Histograms of GDDA values between a single GEO3D graph versus 99 GEO3D *(white)* and GDDA of 30 GEO3D, each versus 99 GEO3D *(grey, the overlap is shown in black.)* (**D**) Histograms of GDDA values between a single ER-DD graph versus 99 ER *(white)* and GDDA of 30 ER, each versus 99 ER *(grey, the overlap is shown in black.)* All networks have approximately the same number of vertices and edges as BG-MS, with graph density of 0.00209. All images were generated by Matlab.

models being inadequate and incommensurable to the network under consideration. Hence, we conclude that neither of the models fit the data.

To verify that our method is indeed capable of classifying a network, we took a GEO3D graph as input and compared it with other GEO3D networks (Fig. 4C). The distribution overlap is clear and the Monte Carlo test gives a P-value of 0.24 ($M = 99$, $N = 99$). Figure 4C also illustrates the possible bias that can occur when just one model graph is used in same model versus model comparison. We emphasize that the graphs used for Figure 4C have the same number of vertices and graph density as BG-MS, and hence they are also in the threshold region, which may account for the relatively low P-value. We also report the GDDA values when one compares an ER-DD query network with ER model networks to show how the method behaves for two closely related models (Fig. 4D). A large overlap between the GDDA values is observed. The P-value for the Monte Carlo test with 100 values ($M = 99$, $N = 99$) is 0.15; hence for a single graph from the ER-DD model, our method cannot reject at the 10% level the (reasonable) null hypothesis that the graph comes from an ER model. Another random graph model designed to model PPIs where edges are drawn between every two vertices according to their degree (Pržulj and Higham, 2006) was also tested with similar results (Supplementary Material); future work will include assessing the fit to other models, such as ER mixture models (Daudin *et al.*, 2008).

In our analysis, we have found that none of the theoretical models considered is suitable for the PPI networks analysed

(see Supplementary Material for the P-values and histograms obtained for the other five PPI networks). However, we provide a statistical framework for comparing real-world networks to other theoretical models using non-parametric statistics.

Our results on GDDA scores suggest that PPI networks are situated in a region of graph density close to the threshold behaviour of the models analysed. *Saccharomyces cerevisiae* has $\sim$6600 protein-coding genes (www.yeastgenome.org) and is predicted to have about 25 000–35 000 interactions (Stumpf *et al.*, 2008); such a network would have a graph density between 0.0011 and 0.0016. For *H.sapiens*, estimates of about 25 000 genes (Human Genome Project) and 650 000 PPI (Stumpf *et al.*, 2008) would also lead to graph densities around 0.002. Both these networks would be placed in the threshold region for the appearance of G_8 as well as G_{17}–G_{27} under the ER model. This may suggest that globally many pathways between proteins are essentially unique, with just a few alternative routes; cliques of size 4 and most graphlets on five vertices are unlikely to appear. Such an architecture would render the network both efficient (not too many edges) and robust (alternative pathways are available).

4 CONCLUSIONS

We have shown that typical values of GDDA, gauged by same model comparison, depend on the number of edges and nodes of the underlying graph.

We propose a statistical method for assessing model fit based on GDDA. Although none of the suggested models fit any of the datasets, we provide the basis for statistical comparison with other models.

The GDDA score is particularly unstable in the graph density region between 0 and 0.01, which encompasses most of the PPI networks currently available. We provide the plausible explanation that this is due to thresholds for the appearance of small subgraphs.

Using these thresholds in ER and GEO3D models as proxy, we suggest that PPI networks themselves tend to operate near the thresholds for the appearance of small subgraphs. That is, the network will start to have a few alternative paths between proteins, but not many. This observation may lead to further conjectures about optimal design of networks, accounting for these critical regimes.

ACKNOWLEDGEMENTS

REFERENCES

Alm,E. and Arkin,A. (2003) Biological networks. *Curr. Opin. Struct. Biol.*, **13**, 193–202.

Alon,N. *et al*. (2008) Biomolecular network motif counting and discovery by color coding. *Bioinformatics*, **24**, 1367–4803.

Bollobás,B. (2001) *Random Graphs*. Cambridge University Press, Cambridge, UK.

Ciriello,G. and Guerra,C. (2008) A review on models and algorithms for motif discovery in protein-protein interaction networks. *Brief. Funct. Genomics Proteomics*, **7**, 147–156.

Costa,L. *et al*. (2007) Characterization of complex networks: A survey of measurements. *Adv. Phys.*, **56**, 167–242.

Dall,J. and Christensen,M. (2002) Random geometric graphs. *Phys. Rev. E*, **66**, 16121–16130.

Daudin,J. *et al*. (2008) A mixture model for random graphs. *Stat. Comput.*, **18**, 173–183.

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

Grochow,J. and Kellis,M. (2007) Network motif discovery using subgraph enumeration and symmetry-breaking. *Lect. Notes Comput. Sci.*, **4453**, 92–106.

Hartwell,L.H. *et al*. (1999) From molecular to modular cell biology. *Nature*, **402** (Suppl. 6761), C47–C52.

Hormozdiari,F. *et al*. (2007) Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. *PLoS Comput. Biol.*, **3**, e118.

Ingram,P. *et al*. (2006) Network motifs: structure does not determine function. *BMC Bioinformatics*, **7**, 1–12.

Ito,T. *et al*. (2000) 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. USA*, **97**, 1143–1147.

Kuchaiev,O. *et al*. (2009) Structure of brain functional networks. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, USA, pp. 4166–4170.

Milenkovic,T. *et al*. (2008) Graphcrunch: a tool for large network analyses. *BMC Bioinformatics*, **9**, 70–81.

Milenkovic,T. *et al*. (2009) Optimized null model for protein structure networks. *PLoS ONE*, **4**, e5967.

Milo,R. *et al*. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.

Penrose,M. (2003) *Random Geometric Graphs*. Oxford University Press, Oxford, UK.

Pržulj,N. and Higham,D. (2006) Modelling protein–protein interaction networks via a stickiness index. *J. R. Soc. Interface*, **3**, 711–716.

Przulj,N. *et al*. (2004) Modeling interactome: scale-free or geometric? *Bioinformatics*, **20**, 3508–3515.

Przulj,N. (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **23**, 177–183.

Przulj,N. (2010) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **26**, 853–854.

Rual,J. *et al*. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, **437**, 1173–1178.

Sharan,R. and Ideker,T. (2006) Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.*, **24**, 427–434.

Shen-Orr,S. *et al*. (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, **31**, 64–68.

Stelzl,U. *et al*. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.

Stumpf,M. *et al*. (2008) Estimating the size of the human interactome. *Proc. Natl Acad. Sci. USA*, **105**, 6959–6964.

von Mering,C. *et al*. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417**, 399–404.