 DOI 10.1002/bies.200900145

When metabolism meets topology: Reconciling metabolite and reaction networks

Raul Montañez,[1] *Miguel Angel Medina,*[1] *Ricard V. Solé,*[2,3] *and Carlos Rodríguez-Caso*[2]*

[1]Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071 Málaga, and CIBER de Enfermedades Raras (CIBERER), Málaga, Spain
[2]ICREA-Complex Systems Lab, Universitat Pompeu Fabra. Parc de Recerca Biomèdica de Barcelona. Dr. Aiguader 88, 08003. Barcelona, Spain
[3]Santa Fe Institute 1399 Hyde Park Road, Santa Fe, NM 87501, USA

The search for a systems-level picture of metabolism as a web of molecular interactions provides a paradigmatic example of how the methods used to characterize a system can bias the interpretation of its functional meaning. Metabolic maps have been analyzed using novel techniques from network theory, revealing some non-trivial, functionally relevant properties. These include a small-world structure and hierarchical modularity. However, as discussed here, some of these properties might actually result from an inappropriate way of defining network interactions. Starting from the so-called bipartite organization of metabolism, where the two meaningful subsets (reactions and metabolites) are considered, most current works use only one of the subsets by means of so-called graph projections. Unfortunately, projected graphs often ignore relevant biological and chemical constraints, thus leading to statistical artifacts. Some of these drawbacks and alternative approaches need to be properly addressed.

Keywords: bipartite networks; hierarchical modularity; metabolic networks; scale-free; small-world

Introduction

More than a thousand chemical reactions occur within our cells, providing the building blocks and the fuel for life. They are linked together forming an intricate, complex network in which metabolites are transformed according to the laws of chemistry and thermodynamics. Metabolism has been traditionally organized in terms of pathways that are interlinked forming a connected roadmap. This roadmap was elucidated through the joint efforts of generations of biochemists during the 20th century. The dispersed information was initially compiled into the famous Boehringer's metabolic map by Gerhard Michal in 1993.[1,2] However, it is only in the last decade that, thanks to new advances in computational methods, metabolic pathways have been conveniently organized and standardized within databases such as KEGG[3] and MetaCyc.[4] Nowadays, metabolic data for a variety of organisms are available.

Metabolism is the best-known cellular network. However, its topological organization – the global pattern of connections of the graph – has not raised the interest of biologists until recently.[5–7] This shift was motivated by the finding that real networks display a number of previously unknown traits such as small-world[8] and scale-free organization.[9,10] Such a picture has permeated a large part of the literature in the field,[11–13] even at the level of standard cell biology textbooks.[14]

The small-world property tells us, roughly speaking, that any two nodes in a network are on average separated by just a few intermediate links, despite the fact that networks are large and sparse. Such short paths could have an immediate impact on network functionality, since they enhance the propagation of changes through the system. Additionally, small-world graphs have a high cliquishness, much longer than expected from random (see Box 1 and Box 2 for a formal definition). Therefore, a small world combines a far-from-random local association with a short-path structure.

The second property is related to the high heterogeneity in the number of links that a given node can display. Specifically, it was observed that the number of nodes having k connections $N(k)$ is such that most of the elements of the net have just one or two links, whereas a handful of them (the hubs) have many connections. This pattern can be described using a well-defined fat-tailed distribution. The presence of hubs has several consequences, from favoring the small-world behavior to inducing internal fragilities associated with their failure. Moreover, the presence of these patterns can be measured, modeled, and used to explore the potential paths followed by these webs through their evolution.[15–19] Network thinking has influenced the study of very diverse systems, from proteomes to the Internet, revealing that very disparate systems share common patterns of organization that can be characterized by their topology.

***Correspondence to:** C. Rodríguez-Caso, ICREA-Complex Systems Lab, Universitat Pompeu Fabra. Parc de Recerca Biomèdica de Barcelona. Dr. Aiguader 88, 08003. Barcelona, Spain.
E-mail: carlos.rodriguez@upf.edu

In a nutshell, a graph can be described as a mathematical abstraction of reality in which nodes are individual units (metabolites of a reaction, species, proteins, actors of a film, or websites) that appear linked by an edge if some type of relation exists among them. In this process, the real system is represented by a connection pattern, thus ignoring most

Box 1

A graph is the mathematical representation of the relation between the elements of a system. Elements are **nodes** or vertices in the graph. Two nodes are connected in the graph by an **link (**or **edge)** if they establish some type of relationship.[61] The number of nodes (N) and edges (E) define the **size** of the graph.

A **bipartite graph** is a graph consisting of two disjoint sets of nodes related through a set of edges. In this graph, connections between nodes of the same set are forbidden. This contrasts to **unipartite graphs**, where elements of a set of nodes establish the connections among them.

The **degree** (k1) or **first neighborhood** of a node is its number of edges or connections. In the bipartite metabolic network, the degree (k1 of a metabolite indicates the number of reactions in which it participates, whereas the degree of a reaction node represents the number of substrates and products of this reaction. **Average degree** $\langle \mathbf{k} \rangle$ is calculated as a global estimator of the network from degree node information. The abundance of the number of nodes of a certain degree is the **degree distribution**. Notice that two $\langle k \rangle$ and two degree distributions can be defined for a bipartite graph. A suitable generalization of the concept of degree for bipartite graphs is just the total number of paths connecting some given node with all nearest-neighboring nodes of the same type, namely **strength**.

The **second neighborhood** (k2) is the number of the neighbors' neighbors of a node. In bipartite metabolic networks, the k2 of a metabolite A indicates the number of metabolites that can be combined in some reaction with A.

Projection of a bipartite graph is the mathematical operation that allows constructing a unipartite graph where nodes of only one type are present. These nodes are connected in the new graph if they share a common neighbor (of the other type) in the bipartite graph. According to this definition, two alternative projections can be done from a bipartite graph (see also Box 2).

A set of nodes forms a **clique** if all possible connections among them are present. Projections of bipartite graphs are enriched of cliques.

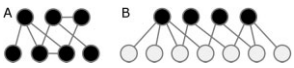
Clustering coefficient is a measure of the local association or "cliquishness" (from clique) of a node. The clustering coefficient of a node is the number of connected neighbors normalized by the number of all possible combinations. **Average clustering** coefficient $\langle C \rangle$ is zero in bipartite graphs. However, it is very high in projections. Calculating the average of clustering coefficient for every k, the **clustering coefficient distribution** is obtained.

A **path** is defined as a sequence of connected nodes. **Minimal path average** is a global estimator of the network. Average is calculated from the minimum of every possible pair of nodes. This measure is closely related to the diameter of a network.

Module is a set of nodes that presents a closer relation, usually defined by a more dense connection, than with the remaining network.

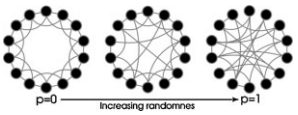
Box 2: Graphs and models

Bipartite graph and projections: For a bipartite graph with two disjoint sets of nodes A and B, the projection of A on B is the unipartite graph constructed with A nodes. In this projection, two A nodes establish a link if they are connected to, at least one, common B node in the bipartite graph (see Fig. 1). Analogously, the projection of B on A can be constructed. For metabolic networks, the two alternative projections of the bipartite graph are known as **metabolite** and **reaction network**.

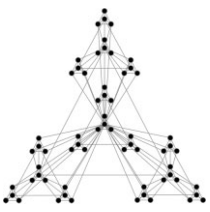


Erdös-Rényi (ER) model[62] **(A)** is a random unipartite graph of N nodes and E edges. The probability P defines the likelihood that a pair of nodes is connected in the graph. ER model follows a Poisson degree distribution where $\langle \mathbf{k} \rangle$ is the mean of distribution. Given N and P, for large enough ER graphs $E \sim P[N(N-1)]/2$, $\langle k \rangle \sim PN$ and $\langle C \rangle \sim P$. Prior to Barabási's work on scale-free networks, it was commonly accepted that the ER model fitted the structure of very large networks. Analogously, the **Erdös-Rényi Bipartite model (B)** can be defined introducing one additional constraint by separating nodes of the network in two groups (labeled in B as white and black nodes). In this network a link between two nodes is established with a probability (P) provided that no pair is established between nodes the same groups. In this network a Poisson degree distribution is given for both sets of nodes and $\langle C \rangle = 0$.

Scale-free graph: This is a network that follows a power-law degree distribution $P(k) \sim k^{-\gamma}$ where gamma ranges between 2 and 3. A handful of nodes (the hubs) have many connections, whereas the vast majority has very few ones. From a theoretical perspective, a consequence of a power-law distribution is that a characteristic scale cannot be defined.



Small-world model: Watts and Strogatz showed in 1999[8] that the small-world phenomenon described by Milgram[63] can be represented by a simple experiment of disordering a regular lattice at random. Rewiring only a handful of links is sufficient to produce a path length similar (L) to that observed in an ER model, but keeping the original local organization of the real network. This local organization is estimated through $\langle C \rangle$. This model allows the small-world criterion for any (unipartite) network to be defined. Real $\langle C \rangle$ and L are compared with the expected values of an ER model with the same number of nodes and links. Formally a graph is small-world when $\langle C \rangle_{(real)} >> \langle C \rangle_{(ER)}$ and $L_{(real)} \sim L_{(ER)}$.



Hierarchical modular model: This is a deterministic model where nodes are clustered in modules in a hierarchical organization. This model, proposed by Ravasz *et al.*[31] presents a $C(k) \sim k^{-1}$. This property has been usually associated in real networks as an indicator of a hierarchical modular organization. Interestingly, this model also exhibits scale-free and small-world behavior.

details in favor of a systemic perspective. Nevertheless, these simplified pictures of reality have proven helpful in providing useful insights into ecology,[20] genomics,[11] neuroscience,[21] or language.[22] However, as shown here, the process of graph construction is an important issue that can bias the conclusions derived from the topological approach.

In this paper we critically examine this view within the context of metabolism. Metabolic maps, along with collaboration,[23–26] mutualistic,[27,28] gene-disease,[29] or drug-target[30] networks belong to the class of so-called bipartite graphs (see Fig. 1 and Box 1 for definitions). Metabolism involves both metabolites and reactions, but metabolic networks have usually been treated as unipartite graphs (networks with only one type of element acting as nodes).[7,31–34] In the seminal works concerning the topological organization of metabolism, such a limitation was overcome by a graph transformation – called projection – of the bipartite network in its unipartite version.[7,31] As summarized in Fig. 1, two alternative graph projections [Pa(G) and Pb(G)] can be obtained from a bipartite graph G. We can construct the so-called metabolite projection in which nodes are metabolites that are linked to each other if they participate in the same reaction (see Box 2). Alternatively, the reaction network is constructed considering nodes as reactions: two reactions are connected if they share a common metabolite. Further work revealed that network projections introduce a strong bias in the results of the topological analysis.[35–38] However, such biases were only addressed within the field of social networks. As shown here, a biologically consistent definition of metabolic network according to its bipartite nature provides a more accurate biological interpretation of its topology.

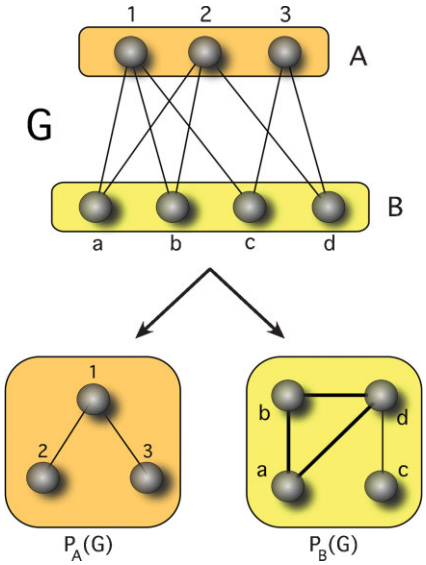


Figure 1. Representation of a bipartite graph. Links establish the connection between the members two disjoint sets: top nodes (A) and bottom nodes (B). Connections between nodes of the same set are forbidden. Two alternative projections (see Box 1 for definition) are possible by considering nodes of one set as connectors of the other one. Notice that triangle connection, a feature measured by the so-called clustering coefficient (see Box 1 for definition) is forbidden in the bipartite graph but not in its projections. Bipartite graph and projections are defined in Boxes 1 and 2.

Metabolism as a complex network: What do we know about metabolic webs?

Metabolism was one of the first real networks identified as both scale-free[6] and small-world[7] with some controversy.[39] Metabolic networks are also considered a paradigmatic example of hierarchical modular organization.[31] Where does this scale-free, modular pattern come from?

Models help us to understand reality and the study of metabolism's organization has not been an exception. Some simple, toy models of reality have provided useful insights into their origin and evolution. An illustrative example of this type of model is the so-called preferential attachment (or rich-gets-richer) mechanism. It has been shown that a graph growing under preferential attachment (see Box 2 for model description) can produce a scale-free network.[40] This simple rule captures the effect of popularity occurring in some real networks such as the Internet[9] or social networks exhibiting scale-free distributions.[10] Another paradigmatic example is given in the paper of Watts and Strogatz, which explains the small-word effect using a very simple network model.[8] This example illustrates the success of a network approximation by giving a very simple explanation about how this pattern can emerge. Finally, hierarchical modularity[31] is another example of a key property displayed by real networks. Roughly speaking, it involves the presence of a nested set of hierarchically assembled modular parts. Such a pattern is reminiscent of fractal structures in nature, and some key properties of real webs can actually be recovered from a fractal-like iterative model (Box 2).

Topological properties are usually evaluated by comparison with null models. The most studied random model is the so-called Erdös-Rényi (ER) graph in which nodes of a set are simply linked by a single probability (see Box 2). The outcome of this model drastically differs from most of the real networks, and it constitutes a reference of what we expect by chance. This model is used to establish the presence of the small-world condition. Analogously, a bipartite version of an ER model is constructed by considering two separated sets of nodes, provided that no connection between nodes of the same group is present (see Box 2). As no other process but randomness is implied in its generation, this model constitutes an excellent null reference for our purposes since it allows us to illustrate the impact of the graph projection' on topological measures.

Interestingly, scale-free, small-world, and hierarchical modularity topological properties match well with the standard view of metabolism from a biological perspective. ATP and $NADH^{+}$ metabolites are examples of hubs (the most connected nodes) in the metabolic network[6] and their topological importance matches their relevant roles as the coins for energetic and reduction power in metabolism, respectively. They participate in most metabolic pathways, whereas many metabolites only take part in single steps of particular pathways, in agreement with the scale-free behavior. The organization of metabolism in coupled reactions and its traditional classification in pathways agrees with the idea of local association and modularity. In this context, cross-links between pathways and the existence of ubiquitous metabolites – the hubs – would contribute to the small-world effect and a hierarchical modular organization.

However, as mentioned above, a network is an abstraction of reality and its construction can determine the conclusions derived from it. At this point, the choice of either a bipartite or unipartite (projected) representation of metabolism acquires special importance. In fact, small world, hierarchy and scale-free features of metabolic networks have been reported using some type of projection. Although these properties seem to be reasonable for metabolism, in light of evidence from other bipartite network studies, the final conclusion is that the topological analysis of metabolic webs should be revisited.

Bipartite nature of metabolic networks and the problem of graph definition

Within metabolic maps, each reaction is not just a causal link between molecules. The flow of matter, kinetic parameters, and time scales can be properly measured or estimated. The global picture is thus rather complete and meaningful. As mentioned above, a graph is an abstraction of reality where only the pattern of connection matters. Figure 2A depicts the standard representation of a metabolic pathway found in any Biochemistry handbook. Although molecular details, enzyme kinetics and most of the complexity of the real system are absent, this representation is enough to sketch the wiring of this particular metabolic process. Thus, a graph based on the metabolic map information is arguably the best choice when looking at how metabolism is connected. However, different graph constructions are possible. In this context, Albert Einstein's famous quote "Make everything as simple as possible, but not simpler'' captures the key point of our work. According to this, a metabolic network should contain the essential information but not less, regarding topological questions, since it provides the clues for a suitable biological interpretation.

Since metabolic networks consist of metabolites that participate in reactions, a special feature of these networks is that metabolites do not interact in pairs, as seen for example with the protein interaction map. In the later, we can use a graph construction based on nodes of the same type interacting in pairs. Metabolic networks require the definition of both reactions and metabolites as separated sets of nodes to reach a meaningful description. In a reaction, a number of substrates (usually more than one) give a number of products. In this case, the connection of metabolites by pairs offers a poorer representation of reality since it cannot properly capture the metabolite association (conversion) in reactions, as bipartite graphs do (see Fig. 2). This bipartite view of metabolism has been commonly addressed in different works.[6,7,41–43]

A number of analyses describing the topological organization of metabolism have introduced substantial simplifications

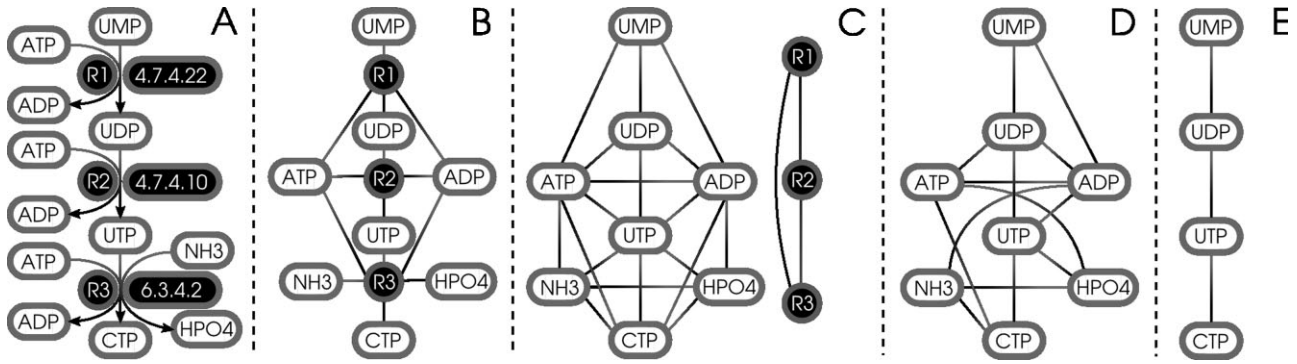


Figure 2. Graph representations of three-step metabolic pathway reproducing the gamma-phosphate transference from ATP to UMP and CTP formation. (A) The standard view of a metabolic pathway in Biochemistry's handbooks. (B) The metabolite-reaction representation in a bipartite graph.[6] (C) Projections of the bipartite graph according to;[7] metabolite projection: all metabolites of a reaction are fully connected (white-node graph in C); and reaction projection: two reactions are connected if they share a common metabolite (black-node graph in C). (D) Alternative metabolite-projection used in ref.[31] Given a reaction, substrates are connected to products, but substrates (and products) are not connected among them. (E) Representation eliminating the ATP and ADP metabolic network hubs applied in ref.[7] and ref.[31] with slight modifications.

to metabolic network definition. In most of them, such simplifications involve the generation of unipartite networks with only metabolites as the nodes of the graph. Figure 2 shows some of these unipartite versions. The standard projection (see Box 1 and Box 2 for definition, and Fig. 2C) is probably one of the most influential network simplifications since it was the first used to define the small-world character of metabolism.[7] In addition, it has been used to describe the power-law distribution of metabolite and reaction unipartite networks.[32]

However, more refined projections have been proposed. A widely used alternative is depicted in Fig. 2D. In this projection the connections occur only between the substrates and products of a reaction.[31,33,44,45] This modification can be justified as a way to capture molecular transformations. Nevertheless, the opposite projection criterion, *i.e.*, the connection between substrates on one side and products on the other (of a reaction), may be similarly justified by the fact that molecular collisions actually occurs between substrates of a reaction. However, as far as we know such a simplification has never been applied.

Additional simplifications in the projection process, such as the elimination of the most connected metabolites due to its ubiquitous nature,[7] or even the elimination of metabolites displaying only one connection and the merging of linear paths in single nodes,[31] have been used for different purposes. In conclusion, a repertoire of unipartite versions of metabolic networks has been used to highlight one particular trait of metabolic organization. However, what determines the correct level of simplification when using projections is an open question. A problem arises when all the pieces of information obtained from different network definitions are put together, ignoring the possibility that such simplifications may introduce a bias.

However, unipartite versions of metabolism are also possible without the use of a projection. An alternative construction of a metabolic network proposed in ref. [39] is probably a fair view of a natural unipartite network. In this representation metabolites are nodes and two metabolites are linked if a carbon transfer occurs between them. In this case, links are not a consequence of a mathematical transformation and they have a physical meaning, even if reactions are not explicit in such a representation. The main limitation of this network is that only a small part of the whole chemical process is considered since not only carbon but other atoms, electrons and energy are actually transferred between molecules. However, extending this idea to any type of transfer would produce a chemically realistic unipartite network. As a side effect, we would lose the information associated with molecular relations due to reactions. However, the combined information of bipartite and this unipartite network would provide a very good insight into the global organization of metabolic topology. Unfortunately, as far as we know, no such network has yet been constructed.

In the next section, we show that simplifications of bipartite metabolic networks produce a loss of information and even generate contradictory results in the case of small-world patterns,[7,39] introducing a bias in the biological interpretation of systemic functional traits.

Too much simple metabolic networks? The problem of graph projection

Wagner and Fell[7] suggested that bipartite graphs are much less intuitive constructs, and less obviously treatable than their projections. It is worth noting that at that time, the analysis of bipartite networks was less developed than for unipartite graphs. Therefore, the topological analysis of the metabolite projection was considered as a reasonable starting point for the study of metabolism as a network.

As we mentioned above, a bipartite network can be transformed in two alternative unipartite versions by means of projection of a graph (see Fig. 1 for a general definition of a graph projection). In metabolite networks, a metabolite projection graph is formed only by metabolites. In this graph, two metabolites are connected by a link if they participate in the same reaction. In contrast, a reaction projection involves reactions as nodes. In this case, two reactions are connected by a link if they share a common metabolite.

As shown in Fig. 3, the same graph projection can be associated with very different bipartite networks. This indicates that only attending to projections, the relation among metabolites through reactions cannot be recovered, and therefore part of the information present in the bipartite graph is lost in the transformation process. Additionally, projections usually exhibit a markedly higher value of average degree in relation to the bipartite graph. As a consequence of projection, average degree has a different interpretation than the bipartite one. As Box 1 indicates, the degree in bipartite graphs has a natural meaning, *i.e.*, the number of partners in a reaction or the number of reactions in which one metabolite participates. In projections as depicted in Fig. 3C, the degree of a node (for example the UTP node) indicates the number of different metabolites that can be related to it. However, compared with the bipartite representation, in such projections we know neither the number of reactions in which UTP is involved nor the number of times that metabolite partners of UTP are repeated in such reactions. Interestingly, the degree of a given node in this (metabolite) projection is recovered by its so-called second neighborhood (see Box 1 for definition) in the bipartite network. Accordingly, it should be stressed that bipartite graphs contain information that cannot be inferred from projections.

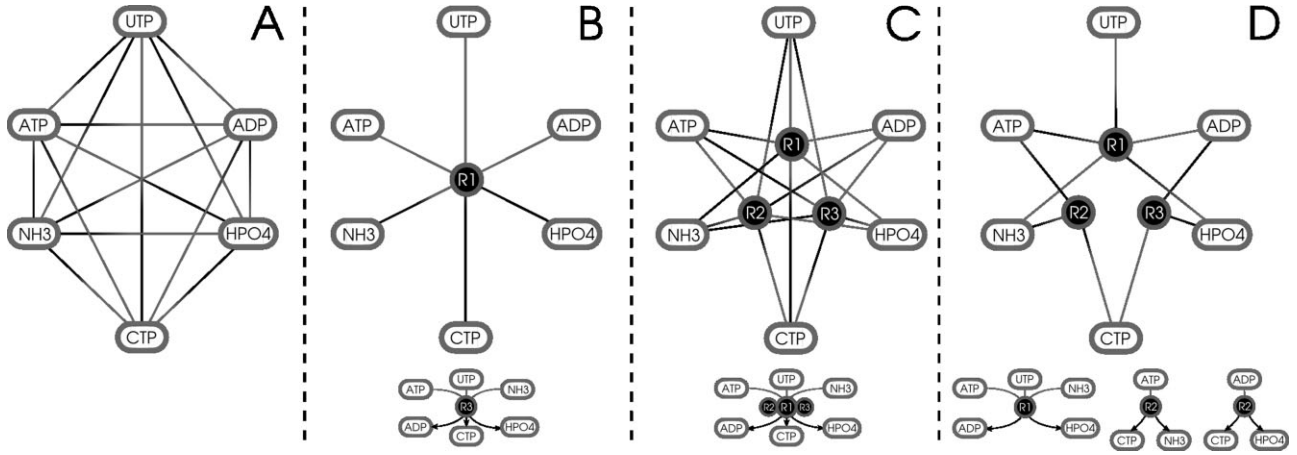


Figure 3. More than one bipartite network can produce the same metabolite projection. (A) Metabolite projection of a three-step pathway representing the gamma-phosphate transference from ATP to CDP. (B) The original bipartite graph with its standard biochemical representation (below). (C, D) Different bipartite graphs originating from non-real reactions that produce metabolic projection like the real one. Notice that a loss of information occurs in the projection.

Another important limitation appears when we look at the degree distribution. Although the degree distribution (see Box 1) for metabolites falls off as a power law,[6] the fact that reactions follow instead a binomial distribution has been largely ignored. To illustrate this, we show in Fig. 4 the degree of distributions for the *E. coli* metabolic network used in ref. [6] in its bipartite version. Figure 4D shows the reaction degree distribution, with an average of 3.73 metabolites and a maximum of 8. This is reasonable since one reaction is a single event, where the number of participants is restricted to a few reactants. We can argue that the minimum number of participants in a reaction is two, corresponding to a tautomeric conversion. However, as Fig. 4D reveals, only a very small fraction of reactions with just one metabolite are present in the network. This is not consistent with the biological interpretation and it may be due to a problem of data curation or incompleteness.

In spite of these small discrepancies, the key problem is that degree distributions have a simple interpretation in the bipartite graph but not in its projections. Similarly, metabolite and reaction distributions are strongly canalized by biological and chemical constraints. In the case of metabolites, the appearance of a new reaction in evolution can be fairly justified by the appearance of new enzymatic activities by a process of gene duplication and diversification. Such a new reaction should occur using existent metabolites and with some probability they may form new compounds that would be incorporated to metabolism. According to this view, a scale-free metabolite distribution is likely to happen and can be indirectly deduced from current unipartite network models of metabolic evolution.[46] However, as reactions mean transformation by molecular collision satisfying both thermodynamic and chemical constraints, the addition or elimination of one molecule to reactions is unlikely to happen. This constraint is more compatible with an exponential/normal distribution than with a scale-free one. This situation has been reported from the study of association networks, where the networks exhibit a scale-free distribution for one node type and normal distributions for the other one. This occurs just because one type of nodes allows a very large range of connections compared to the other.[47,48]

As illustrated in Figs. 1 and 3, a complete picture of the two types of nodes is only captured using bipartite representation, whereas none of the graph projections recover it. In this context, it has been observed that, if one node type follows a power law in the bipartite representation, both projections exhibit power-law degree distributions.[47] This has caused wrong interpretations of the results for reactions, as they have been described to also follow a power-law fashion.[32,49]

Finally, the clustering coefficient (C) is probably the most problematic parameter. Clustering is zero in bipartite networks by definition but not in their projections. An immediate feature derived from the theory is that projections of bipartite graphs are enriched of cliques (see Box 1 for definition) and cliques markedly increase the value of the average clustering ($\langle C \rangle$). For an ER bipartite graph (see Box 2 for description), its projection presents a high $\langle C \rangle$ (see Table 1). However, graph projections do not correspond with a unipartite ER. In fact, $\langle C \rangle$ is much greater in the ER bipartite projection than in an ER unipartite version constructed with the same number of nodes and links of the graph projection (see Table 1 for numerical comparison). This indicates that simple aggregation in a graph induces the clustering after projection. It could be argued that, indeed, clustering is capturing this aggregation,

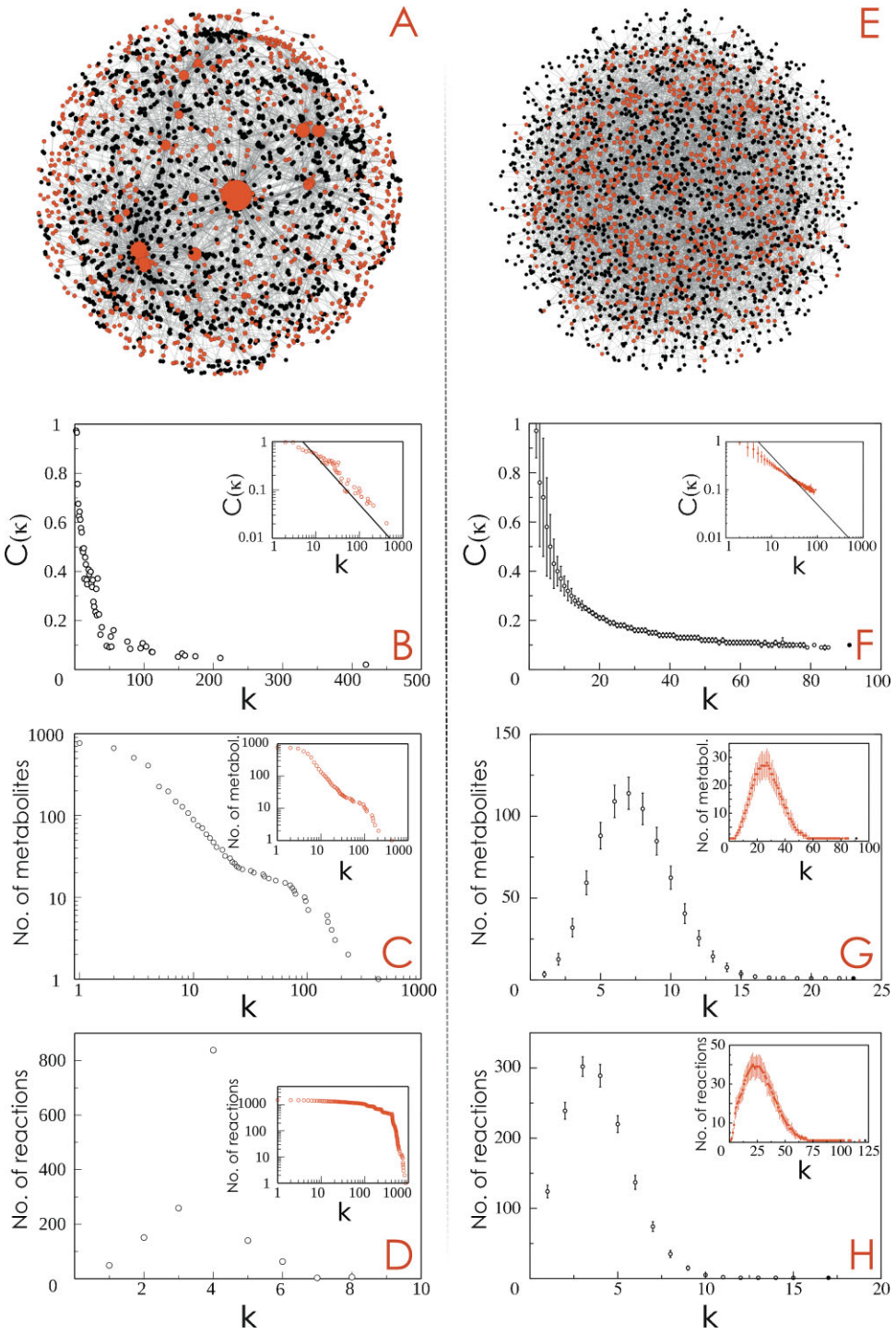


Figure 4. *E. coli* metabolic network, degree, and clustering distributions recalculated from[6] (A–D) and its ER random version (E–H). Both networks present 1509 reactions (labeled in black), 766 metabolites (red), and 5627 links. Node size is proportional to its degree. Power-law decay of $\langle C \rangle$ *versus* k distribution is shown for metabolite projection *E. coli* metabolic network and for its respective ER model as calculated in ref.[31] Logarithmic binning inset (B and F, respectively), reveals the power-law dependence. As a guide for the eye, the k^{-1} slope is indicated in solid line. (C, G) The degree metabolite distribution and respective degree distribution for the metabolite projection inset. (D, H) Degree reaction distribution and degree distribution of the reaction projection inset. ER distributions show the mean and standard deviation of a set of 100 ER graphs.

Table 1. Small-world criterion is satisfied in both *E. coli* metabolic network and in its random null model counterpart. Metabolite and reaction projections of the *E. coli* bipartite network (1), and the respective projections of an ER bipartite model with an identical number of metabolites, reactions, and links of the original network (2), are compared with an ER random graph obtained from the (metabolite or reaction) projection (3). Data from the *E. coli* network construction were obtained from ref.[6] and the resulting network is depicted in Fig. 3. According to the definition, small-world criterion ($\langle C\rangle_{(1 \text{ or } 2)} >> \langle C\rangle_{ER}$ and $L_{(1 \text{ or } 2)} \sim L_{ER}$) is evaluated by comparison with an ER random graph obtained from the (metabolite or reaction) projection (3). In other words, an ER unipartite graph is constructed preserving the same number of nodes and links of the projected network. One hundred of random graphs were generated in each case. Mean and SD for each set were calculated for $\langle C\rangle$ and L values. The results reveal that a real network (1), but also its random bipartite null model (2), are small-worlds when projected. Remarkably, the bipartite nature of a graph is in this case just enough to satisfy the small-world criterion.

	Real network projection (1)		**ER bipartite graph projection (2)**		**ER from projected graph (3)**	
	L	**⟨C⟩**	**L**	**⟨C⟩**	**L**	**⟨C⟩**
Metabolite projection (762 nodes, 5627 links)	2.57	0.67	2.38 ± 0.007	0.19 ± 0.003	3.17 ± 0.003	0.01 ± 0.001
Reaction projection (1506 nodes, 170973 links)	1.82	0.72	2.62 ± 0.006	0.36 ± 0.006	1.84 ± 0.0	0.15 ± 0.001

but actually this gives no additional information to the one already captured by the bipartite graph. An interesting question not yet answered is the likelihood that two metabolites participating in a reaction can also participate in another one. In this case, it seems obvious that clustering measures specifically thought for bipartite networks could contribute to answer this question. In this context, a number of approximations using bipartite graphs has been applied to other real systems[50–53] but not yet translated to metabolism. As previously mentioned, the standard clustering may lead us to wrong interpretations. This becomes more relevant when properties based on clustering such as small-world and hierarchical modularity are evaluated, as seen in the following sections.

Is the metabolic small-world behavior an artifact?

Small-world behavior is a common characteristic of real networks. Its definition requires using both the average of clustering coefficient (C) and minimal path length (L) (see Box 1). For simple (non-bipartite) graphs, the small-world property is often determined by comparing the original graph with a random ER model having the same number of nodes and links.[8,54] However, a problem arises when we deal with bipartite graphs. Since bipartite networks do not exhibit clustering, the presence of a small-world behavior is here meaningless and only unipartite networks can be properly evaluated. Therefore, a projection is required. Once the projected network is obtained, it can be compared with its associated graph obtained from the ER model. Notice that such an ER model has the same number than nodes and links of the projected network. But, what happens if we do the same with the ER bipartite graph?

As mentioned above, clustering appears as a result of the projection process and this also occurs when ER bipartite graphs are projected. According to this, to establish a small-world criterion we must compare this network with its respective ER model derived from the projected bipartite ER graph. A remarkable fact is that the projection of an ER bipartite network, roughly satisfies a small-world pattern (see Table 1 for a numerical example). Some questions now arise: can we ensure that metabolic or even any bipartite graph is small world when its random bipartite version also satisfies this condition? Is the simple aggregation by reactions an appropriate requisite to conclude that metabolism is small world, or is it a side effect of the graph construction algorithm?

As Table 1 indicates, the *E. coli* metabolite projection reveals a small-world behavior but also its ER bipartite correlates when they were projected as described in Fig. 2C (see also ref.[7]). According to the arguments provided here, we cannot guarantee that the small world of metabolic networks has a biologically meaningful interpretation (such as an optimization of the metabolic flux) since a null bipartite model also fulfills such a condition. On the other hand, we cannot reject the idea that metabolisms are organized within the small-world constraints. Clustering of real networks is greater than that of their bipartite random null model counterparts (Table 1), thus providing indirect evidence of some local association of metabolism beyond a simple aggregation effect. However, as small world is a qualitative behavior, such a comparison cannot be properly done.

Alternative projections have been described, such as that displayed in Fig. 2D, that also share small-world behavior.[33,44,45] However, the impact of such a projection on clustering has not yet been explored in depth. The main limitation is that it requires additional considerations beyond the standard ER bipartite model, such the estimation of what number of substrates and products are expected by chance

for every reaction. Further research in this direction would contribute to addressing our open questions.

Summarizing, the main problem is that the standard definition of small world is not suitable for bipartite networks and, therefore, the search for a more convenient small-world criterion for these graphs may provide an answer as to whether metabolic networks are really small worlds. According to this idea, some clustering coefficient definitions have been proposed for bipartite graphs, but little of its relevance for the small-world property has been explored.[52,53] It is worth stressing that other arguments, such as modularity, go in favor of the local association required for small-worldness but as seen below, the effect of hierarchical modularity also may be affected by a wrong interpretation of clustering in the graph projections.

Do metabolic networks exhibit hierarchical modularity?

It has been suggested (and widely used) that an indicator of a hierarchical organization is the power-law dependence of $\langle C \rangle$ *versus* k.[31,37,49,55] Clustering captures the cliquishness of a node (measuring the probability that two nodes sharing a common neighbor are also connected among them). It was found that degree and clustering are negatively correlated: nodes with low k are more likely to form clusters than those with high degree. This idea is captured by a toy model of graph generation. This is illustrated in Box 2 where a complex, fractal-like network is generated by iteratively reconnecting an initial subgraph. As this model is far from realistic, it has been used as a metaphor for complex networks displaying modular and hierarchical scale-free organization. The topological analysis of this web reveals a C(k) decay, namely:

$$C(k) \sim k^{-\alpha}$$

with $\alpha = 1$. Such a scaling law has been observed in a number of real systems, although with different scaling exponents,[56] and this power-law dependence has been widely adopted as an indicative for hierarchy. In addition, it has been suggested that scale-free behavior is necessary, but not sufficient, condition to produce such scaling behavior.[55] Actually, metabolic networks (or rather metabolite projections) exhibit such a scale-freeness and a clustering decay close to that predicted by the model. Although graph construction can enhance clustering, we show in Fig. 4B that a standard metabolite projection (as shown in Fig. 2B) is quite close to a power-law decay with $\alpha = 1$.

Considering the effect of graph projection on clustering, the following question has to be formulated. Does projection introduce some bias in the C(k) distribution? As with the small-world pattern, Fig. 4F indicates that projection is just enough to produce a power-law dependence on C(k). It is worth noting that such a decay in the projected ER bipartite model is not so pronounced as in the real (projected) network (see Figs. 4B and F). Actually, such a dependence occurs in the ER bipartite null model even without satisfying the scale-free condition and being dramatically distant from the prototypical hierarchical one (see Box 2). Is the hierarchical criterion correct? Beyond this general question, since we cannot establish the contribution of projection on clustering decay, we cannot properly talk about hierarchy in metabolic networks. Consequently, a criterion for hierarchy avoiding the use of projections is required and further research in this direction would contribute to clarifying whether metabolic networks are truly hierarchical.

Similarly, modular measures, particularly those based on clustering coefficient measure,[31,57,58] may be affected by the projection effect. In this context, modularity detection in bipartite graphs would contribute to clarifying the modular character of metabolism. Methods to detect modularity in bipartite networks have been recently published,[59,60] representing a promising alternative to previous approaches.

The bipartite network approach

As discussed above, some well-established measures for bipartite graphs offer an ideal framework for the study of metabolism, keeping the relation between metabolites and reactions.

An alternative to commonly applied measures in unipartite graphs is the 'strength'[48] (see Box 1 for definition). In the case of metabolism, it is possible to figure out how important a metabolite can be through its degree. However, one can imagine that, for instance, not only the ATP hub but also the ATP-ADP pair is biologically relevant, since it represents a pair transformation shared by large number of reactions. This information is captured by the strength, and its biological interpretation is straightforward. As metabolism occurs by transformations, strength would indicate the importance of a substrate-product pair due to its participation in a large number of reactions. However, in spite of the potential of strength, it has not been applied to metabolism so far.

Strength and the use of a suitable version of bipartite clustering could provide new insights on metabolic organization. In addition, a number of properties such as modularity or small-world behavior should be revisited according to a bipartite view, avoiding the undesirable effects associated with projections.

In addition, the bipartite representation offers a better picture of metabolism since direction and stoichiometry can be included to produce a more reliable view of metabolism. The study of directed bipartite paths, cycle detection, and degree correlations represent still unexplored ways in the study of metabolic maps.

Conclusions and perspectives

Topological studies of metabolism have been developed over the last few years by using graph representations of pathways. In this work we have shown that a network description imposes strong constraints to our abstraction of reality. In the case of metabolism, the use of unipartite versions can lead to wrong interpretations of some of the most relevant graph attributes. Average degree and degree distributions are paradigmatic examples of this problem. Instead, the bipartite view offers a cleaner interpretation of topological features. A note of caution is thus needed in relation to the effect of projection on clustering and how this produces some properties "for free", such as small-world structure and hierarchy. Finding universalities in complex networks is an intriguing and fascinating issue. However, this is meaningful only when real systems are conveniently mapped into a network abstraction. In this context, the more natural representation of metabolism as a bipartite graph and the development of appropriate measures based on such a representation are much needed.

Acknowledgments: This work was supported by the EU 6th framework project ComplexDis (NEST-043241, CRC), James S. McDonnell Foundation and Santa Fe Institute (RVS), grants SAF 2008-02522 (Spanish Ministry of Science and Innovation), Fundación Ramón Areces, P07-CVI-02999 and group BIO-267 (Andalusian Government). The "CIBER de Enfermedades Raras" is an initiative of the ISCIII (Spain). The authors would like to thank Bernat Corominas-Murtra, Francisca Sánchez-Jiménez, and Alejandro Real-Chicharro for useful comments.

References

1. **Michal G.** 1993. Biochemical Pathways (poster). Boehringer Mannheim: Penzberg.
2. **Michal G.** 1998. On representation of metabolic pathways. *Biosystems* **47**: 1–7.
3. **Kanehisa M**, **Araki M**, **Goto S**, *et al.* 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36**: D480–84.
4. **Caspi R**, **Foerster H**, **Fulcher CA**, *et al.* 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **36**: 623–31.
5. **Ouzounis CA**, **Karp PD.** 2000. Global properties of the metabolic map of *Escherichia coli*. *Genome Res* **10**: 568–76.
6. **Jeong H**, **Tombor B**, **Albert R**, *et al.* 2000. The large-scale organization of metabolic networks. *Nature* Oct 5; **407**: 651–54.
7. **Wagner A**, **Fell DA.** 2001. The small world inside large metabolic networks. *Proc Biol Sci* **268**: 1803–10.
8. **Watts DJ**, **Strogatz SH.** 1998. Collective dynamics of 'small-world' networks. *Nature* **393**: 440–42.
9. **Albert R**, **Jeong H**, **Barabási AL.** 1999. Internet: diameter of the world-wide web. *Nature* **401**: 130–31.
10. **Barabási AL.** 2009. Scale-free networks: a decade and beyond. *Science* **325**: 412–13.
11. **Koonin EV**, **Wolf YI**, **Karev GP.** 2006. Power Laws, Scale-Free Networks and Genome Biology (Molecular Biology Intelligence Unit). Berlin: Springer.
12. **Csermely P.** 2006. Weak links. Stabilizers of complex systems from proteins to social networks. Berlin: Springer.
13. **Képès F.** 2007. Biological networks, Complex systems and interdisciplinary science Vol. **3**. New Jersey: World Scientific.
14. **Alberts B**, **Johnson A**, **Lewis J**, *et al.* 2002. Molecular Biology of the Cell, 4th edn. New York: Garland Science.
15. **Wagner A.** 2002. Asymmetric functional divergence of duplicate genes. *Mol Biol Evol* **19**: 1760–68.
16. **Solé RV**, **Pastor-Satorras R**, **Smith ED**, *et al.* 2002. A model of large-scale proteome evolution. *Adv Complex Syst.* **5**: 43–54.
17. **Rodríguez-Caso C**, **Medina MA**, **Solé RV.** 2005. Topology, tinkering and evolution of the human transcription factor network. *FEBS J* **272**: 6423–34.
18. **Rodríguez-Caso C**, **Corominas-Murtra B**, **Solé RV.** 2009. On the basic computational structure of gene regulatory networks. *Mol Biosyst* **5**: 1617–29.
19. **Knight CG**, **Pinney JW.** 2009. Making the right connections: biological networks in the light of evolution. *BioEssays* **31**: 1080–90.
20. **Montoya JM**, **Pimm SL**, **Solé RV.** 2006. Ecological networks and their fragility. *Nature* **442**: 259–64.
21. **Bullmore E**, **Sporns O.** 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* Mar; **10**: 186–98.
22. **Solé RV**, **Corminas-Murtra B**, **Valverde S**, *et al.* Language networks: their structure, function and evolution. *Complexity* (in press).
23. **Mariolis P.** 1975. Interlocking directorates and control of corporations: the theory of bank control. *Soc Sci Q* **56**: 425–39.
24. **Newman MEJ.** 2001a. Who is the best connected scientist? A study of scientific coauthorship networks. *Phys Rev E* **64**: 016131.
25. **Newman MEJ.** 2001b. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys Rev E* **64**: 016132.
26. **Davis GF**, **Yoo M**, **Baker WE.** 2001. The small world of the corporate elite. New Orleans: Social Networks Sunbelt XXI Conference.
27. **Bascompte J**, **Jordano P.** 2007. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu Rev Ecol Evol Syst* **38**: 567–93.
28. **Rezende E**, **Lavabre J**, **Guimaraes P**, *et al.* 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**: 925–8.
29. **Goh KI**, **Cusick ME**, **Valle D**, *et al.* 2007. The human disease network. *Proc Natl Acad Sci U S A* **104**: 8685–90.
30. **Yildirim MA**, **Goh KI**, **Cusick ME**, *et al.* 2007. Drug-target network. *Nat Biotechnol* Oct; **25**: 1119–26.
31. **Ravasz E**, **Somera AL**, **Mongru DA**, *et al.* 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**: 1551–5.
32. **Burgos E**, **Ceva H**, **Hernández L**, *et al.* 2008. Two classes of bipartite networks: nested biological and social systems. *Phys Rev E* **78**: 046113.
33. **Ma HW**, **Zeng AP.** 2003. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**: 1423–30.
34. **Bilke S**, **Peterson C.** 2001. Topological properties of citation and metabolic networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **64**: 036106.
35. **Newman ME**, **Watts DJ**, **Strogatz SH.** 2002. Random graph models of social networks. *Proc Natl Acad Sci U S A* **99**: 2566–72.
36. **Newman ME**, **Park J.** 2003. Why social networks are different from other types of networks. *Phys Rev E* **68**: 036122.
37. **Vázquez A.** 2003. Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Phys Rev E* **67**: 056104.
38. **Ramasco JJ**, **Dorogovtsev SN**, **Pastor-Satorras R.** 2004. Self-organization of collaboration networks. *Phys Rev E* **70**: 036106.
39. **Arita M.** 2004. The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci U S A* **101**: 1543–7.
40. **Barabási AL**, **Albert R.** 1999. Emergence of scaling in random networks. *Science* **286**: 509–12.
41. **van Helden J**, **Wernisch L**, **Gilbert D**, *et al.* 2002. Graph-based analysis of metabolic networks. *Ernst Schering Res Found Workshop* 245–74.
42. **Zhou T**, **Ren J**, **Medo M**, *et al.* 2007. Bipartite network projection and personal recommendation. *Phys Rev E* **76**: 046115.

43. **Lacroix V**, **Cottret L**, **Thébault P**, *et al.* 2008. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans Comput Biol Bioinform* **5**: 594–617.
44. **Zhao J**, **Yu H**, **Luo JH**, *et al.* 2006. Hierarchical modularity of nested bowties in metabolic networks. *BMC Bioinformatics* **7**: 386.
45. **Szymanski J**, **Jozefczuk S**, **Nikoloski Z**, *et al.* 2009. Stability of metabolic correlations under changing environmental conditions in *Escherichia coli* – A systems approach. *PLoS One* **4**: e7441.
46. **Light S**, **Kraulis P**, **Elofsson A.** 2005. Preferential attachment in the evolution of metabolic networks. *BMC Genomics* Nov 10; **6**: 159.
47. **Guillaume JL**, **Latapy M.** 2006. Bipartite graphs as models of complex networks. *Physica A* **371**: 795–813.
48. **Tanaka R.** 2005. Scale-rich metabolic networks. *Phys Rev Lett* **94**: 168101.
49. **Albert R.** 2005. Scale-free networks in cell biology. *J Cell Sci* **118**: 4947–57.
50. **Borgatti SP**, **Everett MG.** 1997. Network analysis of 2-mode data. *Soc Networks* **19**: 243–69.
51. **Robins G**, **Alexander M.** 2004. Small worlds among interlocking directors: network structure and distance in bipartite graphs. *Comput Math Organization Theory* **10**: 69–94.
52. **Lind PG**, **González MC**, **Herrmann HJ.** 2005. Cycles and clustering in bipartite networks. *Phys Rev E* **72**: 056127.
53. **Zhang P**, **Wang J**, **Li X**, *et al.* 2008. Clustering coefficient and community structure of bipartite networks. *Physica A* **387**: 6869–75.
54. **Dorogovtsev S**, **Mendes J.** 2003. Evolution of networks: From biological nets to the Internet and WWW. Oxford: Oxford University Press.
55. **Barabási AL**, **Oltvai ZN.** 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet* Feb; **5**: 101–13.
56. **Ravasz E**, **Barabasi AL.** 2003. Hierarchical organization in complex networks. *Phys Rev E* **67**: 026112.
57. **Clauset A**, **Newman MEJ**, **Moore C.** 2004. Finding community structure in very large networks. *Phys Rev E* **70**: 066111.
58. **Yip AM**, **Horvath S.** 2007. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* **8**: 22.
59. **Guimerà R**, **Sales-Pardo M**, **Amaral LA.** 2007. Module identification in bipartite and directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **76**: 036102.
60. **Lehmann S**, **Schwartz M**, **Hansen LK.** 2008. Biclique communities. *Phys Rev E* **78**: 016108.
61. **Bollobás B.** 1979. Graph theory: an introductory course. New York: Springer Verlag.
62. **Erdös P**, **Rényi A.** 1959. On random graphs. *I Publ Math* **6**: 290.
63. **Milgram S.** 1967. The small-world problem. *Psycol Today* **2**: 60–7.