Networks Regulating Networks: The Effects of Constraints on Topological Evolution

Francisco C. Santos[1], Hugues Bersini[1], and Tom Lenaerts[2,3]

[1] IRIDIA, CoDE, Université Libre de Bruxelles, Brussels, Belgium[**]
[2] SWITCH Laboratory, VIB, Brussels, Belgium
[3] Vrije Universiteit Brussel, Brussels, Belgium

Abstract. We propose a generalized framework to analyse constraints and representations in growing complex networks. We show that the introduction of biological, social and technological information by means of an additional network of constraints, together with the distinction between complete potential networks and instantaneous effective ones, can offer additional insights about the final topological outcome. Specifically, we study the emergence of exponential cutoffs in broad-scale degree distributions as a result of high level constraints.

1 Introduction

In the last years, it has been experimentally observed that real-world biological, social and technological networks are not structured in a random way [1]. Instead, most of these networks are organized in such a way that a few nodes are able to interact with a lot whereas many others only interact with a few. The extreme case is often referred as scale-free networks in which the degree distribution follows a power-law [2]. However, other configurations showing lower levels of heterogeneity are also common [3]. Dissections of real world networks have produced evidence for single-scale networks, characterized by a fast Gaussian decaying tail in the degree distribution, broad-scale distributions, defined by a power-law with an abrupt truncation for large connectivities [4] and, finally, the previously referred scale-free class. The ubiquity of such classes of networks raises many questions of which one is on the origin of these topological properties. Moreover, in the context of understanding complex social and biological phenomena, it is useful, and in many cases mandatory, to understand the topology of the underlying networks of interactions [1,5,6]. System's global properties rely extensively upon the underlying network and different dynamical outcomes emerge from different topologies [7,8,9,10]. One way of understanding these topological features relies on the development and the study (both mathematically and in software) of abstract models showing the same degree distribution as observed in empirical observations [2,11,5,12,13,14]. Although these universal mechanisms provides rules on how to construct networks that belong to the different classes,

---

[**] For further information contact fsantos@ulb.ac.be
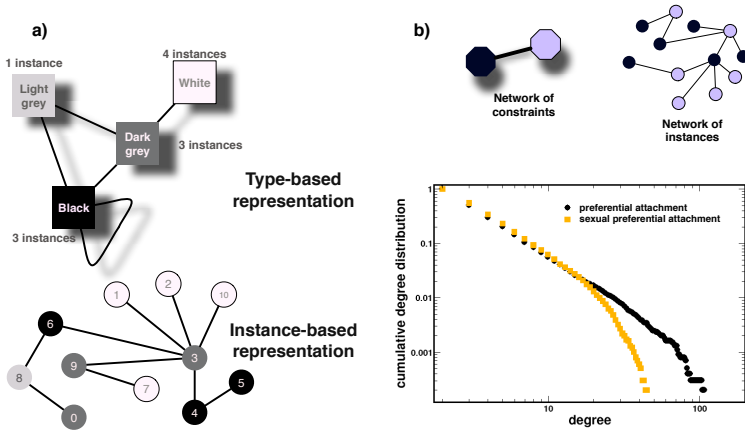
Fig. 1. a) Instance and Types. Instance-based: each node represents an instance of a certain type represented by a different colour (ex: social and technological networks). Type-based: each node represents a certain type with concentration equal to the number of instances of that type. (ex: biological and chemical networks). b) Sexual preferential attachment. Influence of intrinsic features of each node in the final degree distribution. Black dots stand for the cumulative degree distribution after growing a network of 1000 nodes using the Barabási and Albert Model of growth and preferential attachment [2] and the squares show the resulting cumulative degree distribution of the same model but constraining the preferential attachment rule to nodes of the opposite sexual tag. Here, we implicitly define a network of constraints with two nodes (without loops) and one link.

they do not clarify the domain-specific rules that explain the network topology in terms of the specific social, biological or technological domain. Since this explanation is essential for researchers in those particular domains, the models miss some key properties of the networks they aim at replicating. An incomplete set of the constructive laws that form the network can only lead to an incomplete assessment of the empirical data and of the history that produces them. This article proposes an enrichment of the modelling vocabulary and practice used to study growing networks. First we argue that there exists for each network a set of constraints that define its topology. Knowing this type restrictions will provide insight in how generalised growth functions can be defined. Second, we discuss that this set of type constraints define the network of all possibilities which is not necessarily the same as the network that is derived from the data. This differentiation between the potential and the instantaneous network underlines our argument that fitting generalized models to the data is not the best strategy since it does not explain the underlying rules of formation. To illustrate our argument we provide a few examples. Note that the examples are hence only purely illustrative. The main goal here is introduce a slightly richer common language to characterise complex networks and their evolution in time.

2 Defining Nodes

In biology, mean-field analysis is often used to attain an understanding of the complex dynamical system common in natural systems. For instance, in biochemical reaction networks, nodes are often associated to chemical components and links to chemical reactions [9]. A similar approach is often taken in the study of protein interaction networks [15,16]. A node is assumed to represent a certain protein and a link is drawn if binding is empirically observed between two proteins. Since each individual element, at the atomic or molecular level, is clustered in a node representing a certain type, a concentration value is often associated with the node. Similar to the two networks shown in Figure 1-a, nodes are types with particular physical properties, and a concentration value informs about the number of instances of any particular type. A similar type/instance dichotomy can be found in other biological networks such as food web networks, signal transduction networks, etc. As a rule, biological networks are always represented as type-based networks. There are two main motivations for this. First, most of the experimental data can be obtained only in this format. Second, it would almost be impossible to define such kind of networks at an instance level. Either it becomes trivial (for instance, it is hard to imagine a lion interacting with more than one specie at a time t) or too complex. When an instance of a protein interacts with another one, it will become part of a complex and, as a result, won't be able to connect to any other instance. In addition, the intrinsic properties of such node would no longer be the same. The type-level is indeed the obvious level of representation for biological networks of interactions.

Is this type/instance dichotomy meaningful in other scientific domains? It seems not. Social and technological networks are often studied by adopting another perspective. Here every node corresponds to one particular person, computer or HTML page, to give the three most usual examples. The justification of such a choice is mainly practical. First, this is the way in which empirical results are obtained. Second, in most of the cases, the consideration of type is simply meaningless. Defining a type-based network entails a lost in the amount of information (see Figure 1-a). Nonetheless, types are always there, and for each case a proper type representation could be found.

In biology, most of the recent modelling of network evolution does not take this distinction into account. First, types are treated as instances in order to drop the cumbersome issue raised by the dynamics at the level of the number of instances/concentration. This might be the best approach for both social and technological networks but is clearly oversimplified for chemical or biological ones. The absence of any biological or chemical information inherent in the definition of types leaves out some physical constraints certainly decisive in the network evolution. For instance, some nodes interact because their physical characteristics make such interactions extremely likely. This is definitely not related with any kind of topological consideration (like the amount of links that each of agents has). Accordingly, it is hard to accept a system that grows following a degree-preferential rule as biologically realistic. Some highly connected nodes are so, because they have intrinsic features that allow them to connect to a higher

number of partners. Roughly, they are natural hubs or *born to be hubs* instead of being hubs just because of their presence in the network during a longer period. All these special features are included in the definition of each type and in the way links are defined. By disregarding concentration and the number of instances of a certain type, we loose one of the possible mechanisms behind the topological evolution of a growing network. The dynamics of the concentration/instance level often produces the necessary feedback for the network growing while this growing in turn constrains the dynamics. This constant interplay between the dynamics and the topological dynamics is something not taken into account in most of the network modelling.

3   Constraint Ties Among Networks

Adding a natural semantic to the definition of nodes entails the presence of an additional level of representation (we will call it *level 1* in the following) that will constrain what happens at the lower level (called *level 0* in the following). This is true in all kinds of networks. Some nodes, simply are not able to connect to some other nodes. These additional constraints may be due to all sort of reasons. For example, in social or economical networks, a European is more likely to interact with other Europeans. Considering such constraint tags when modelling the growth of networks, different results will be produced as a function of these constraints and their strength. Let's consider a first very simple example of sexual interaction network, where links and nodes represent sexual interactions and individuals, respectively. Such an interaction network has been shown to follow a scale-free degree distribution or at least to be part of a broad-scale class of interaction networks [17,18]. The evolution of this network is successfully modelled by a preferential attachment and growth model, proposed in [2]. In this case, an obvious constraint can be introduced in order to restrain the study to heterosexual relationships. This additional constraints network (*level 1* in Table 1) is simply defined by two gender nodes (male and female) connected by a link but without any loop (that would represent homosexual connections). Considering growth and preferential attachment model and stating that each node has 50 per cent chance of being part of one of the genders, the preferential attachment process will be restricted to the subset of nodes belonging to the opposite gender. The final result reflects precisely that (see Figure 1-b). The heterosexual constraint introduces an artificial cut-off that is frequently observed in experimentally derived networks [3].This sexual network example illustrates how constraints, reflecting intrinsic features of each node, influence the final outcome of the network. As discussed in the following paragraphs, more complex constraint networks will produce more complex effects.

In biology, proteins and chemical nodes have often been modelled by a key-lock mechanism [19,20,21,22,14]. It is assumed that nodes are defined by bit-strings and links by a pre-defined affinity function as in [14]. In this framework, the hamming distance between 2 bit-strings and a global or individual connection threshold have been classically used as affinity function, but many more choices
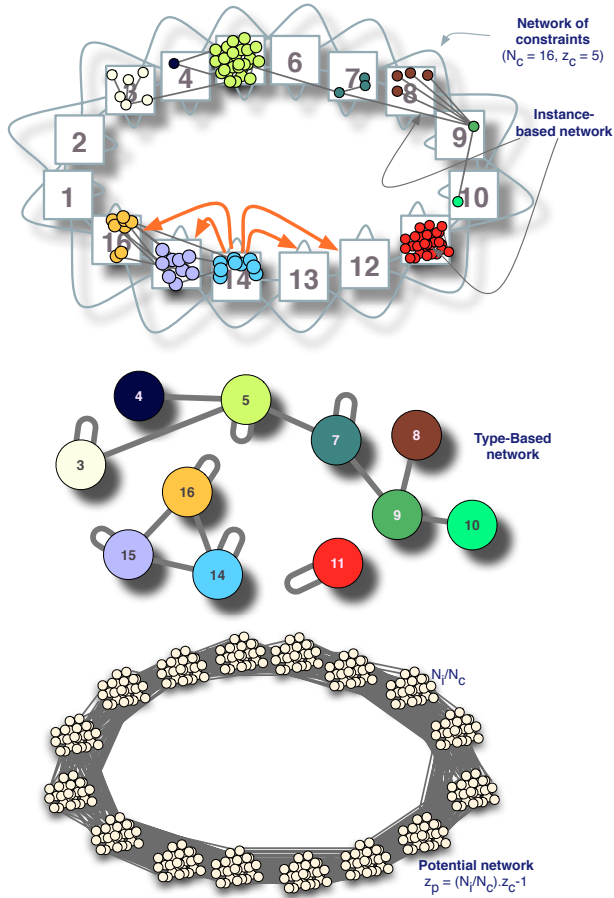
Fig. 2. Spatial constraints at different levels of representation. Upper panel: we have the network of constraints (squares and light grey links) with N_c nodes (16 in this example) and z_c as average degree (we allow loops at the constraints level, allowing instances of the same type to potentially interact). Each number/colour represents a different type. On top of the the constraints network, we can observe an example of an effective instance-based network, with $N_i = 42$ nodes. This last network helps to understand the correspondent type-based network in the middle panel, emerging as a subset of the constraints network. Finally, in the last panel, we see a potential network, drawn from the constraints network, with a maximum number of N_p nodes and z_p links per node. Each node will be part of a certain type, and will connect to all members of its type and to the ones belonging to types that have a link (at the constraints level) with its own type.

Table 1. Two basic distinctions: level 0 *vs* level 1 and instantaneous *vs* potential. We characterize as instantaneous the effective network at *level 0* and individual types at *level 1*. They are the interactions spotted at a given instant of time and turn out to samples of all interactions defined as potential ones. The instantaneous network, either at *level 0* or *1*, is a subset of nodes and links of the potential one originating from the structural evolution of the network.

	Level 0	Level 1
Potential	Potential network	Network of constraints
Instantaneous	Instance-based Representation	Type-Level Representation
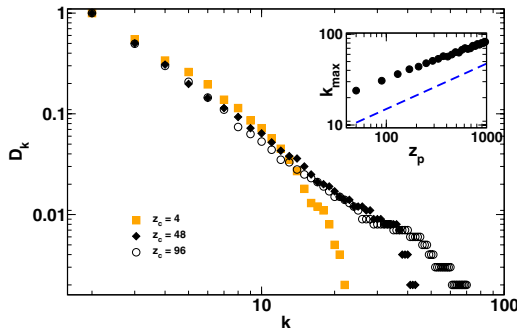
Fig. 3. Cumulative degree distribution (D_k) for different values of average connectivity at the constraints level (z_c). Inset: Maximum degree (k_{max}) dependence on the average degree at the potential level, $z_p = z_c \frac{N_i}{N_c} - 1$, where N_i and N_c represent the number of instances and the number of types, respectively. The dashed line has slope $1/2$ and indicates the expected relationship between z_p and the finite-size cut-off given by $q_c \sim N^{\frac{1}{1-\gamma}}$ (see main text).

are possible. As a matter of fact, defining nodes with a specific tag or intrinsic characteristics, together with a list or a network of all constraints implicitly gives rise to a set of all possible interactions that can be represented as a potential network at either *level 0* or *level 1*. The effective or observed network, that will be called in the following *the instantaneous network*, boils down to a subset of nodes and links of this potential scenario. In the Barabási and Albert (BA) model of growth and preferential attachment [23], no constraint is defined and there is no need for an extra level of representation. The potential network at *level 0* is just a fully connected graph of size N and average degree $z_p = N - 1$. On the other hand, in the BA model with sexual constraints described before, the *level 1* potential network is simply a two nodes network with no loop. The *level 0* potential network turns out to be a fully connected bi-partide graph, reducing the number of potential links to $z_p = \frac{N}{2}$ inducing a finite-size effect [11] in the maximum connectivity of network (see below).

In the previous paragraphs, we introduced different representations and ways of constraining the network growth. Table 1 attempts at summarising and clarifying these distinctions. At the lowest level, *level 0*, first appears the instantaneous network of interactions between the individual instances currently present in the network. It is the real *physical* level at which the interactions take place. From that one, we can derive the instantaneous network at *level 1* reflecting the current constrained interactions between each type or flag that characterises all nodes present at *level 0*. It is impossible to obtain the individual interactions among instances from the network established at *level 1*, while the other way is the only possibility. Finally, all possible interactions at these two levels of representation are defined by a potential network, both at *level 0* and *level 1*. While the instance-based instantaneous network represents a time snapshot of the potential network, similarly the type-based instantaneous network represents a subset of the potential network of constraints. Some nodes likely to be present in the instantaneous network may not actually be there. In the sexual network example, the *level 1* potential network coincides with the *level 1* instantaneous network, since both genders are present at the instance level. This is not a general feature, though. One can easily devise a model where some of the types are not present in the instantaneous *level 0* network, reducing the size of the *level 1* instantaneous network to a very small part of the potential one.

By extending the previous trivial network of gender constraints, a similar model of growth and preferential attachment can be imagined, but with many more types and constraints. Let's start by introducing at *level 1* a general graph of constraints of size N_c and average degree z_c. Only instances belonging to connected types are able to interact. We can construe this network as describing possible interactions among families of proteins, international economical transactions (where each type is a country), or even shaped by religious or spatial constraints. Each of these examples gives rise to a different network of constraints with non-trivial topologies. A network with spatial constraints is certainly the easiest to illustrate. It can be characterized, in its simplest form, by a regular network [24]. Figure 2 shows how all four kinds of networks (*level 0* and *level 1*, *potential* and *instantaneous*) influence each other. If we consider that every node of type j can interact with all nodes present in a same square and with all nodes present into squares in the neighbourhood (see Figure 2), the potential network is (locally) a fully-connected network among all nodes of the same square and among all nodes of squares which are positioned side by side. Once again, the instantaneous networks (at both *levels 0* and 1) may differ from the potential network, regarding the present number of links and nodes, but will always be a subset of it.

Figure 3 shows the simple implementation of this idea by a generalised constraints network, where the effect of a regular network of constraints at *level 1* on the final structure of the network at *level 0* is studied. A regular network of constraints of size N_c (10^2 types) is considered, with an average degree $z_c = 5$, and $N_i = 10^3$ instances (a subset of all possible instances N_p) uniformly spread among all types. At each time step, a new instance is added and connect preferen-

tially [2] to highly connected instances already present in the network. Depending on the average degree (z_c) of the potential network at *level 1*, a different average connectivity will be produced for the potential network at *level 0* (z_p). Each instance will be restrained to interact with the elements of its own type or flag and also with the instances belonging to the neighbourhood types. To simplify, we will consider in the following a finite potential network, in which the number of instances N_i has reached the maximum number of possible instances N_p. Hence, the potential average degree at *level 0* will be given by $z_p = z_c \frac{N_i}{N_c} - 1$, where N_i represent the number of instances and N_c the number of types. The fact that the average degree of the potential network is no longer equal to $N_i - 1$ (like in the usual Barabási and Albert model [23] where every node can potentially connect to every other node), induces typical finite-size effects. The preferential attachment universe is now reduced to the universe of potential partners (z_p), and the finite-size cut-off that scales up proportionally to $N^{\frac{1}{1-\gamma}}$ ($\gamma = 3$ being the resultant slope of the degree distribution $d_k = k^{-\gamma}$) [11] instead scales up with $z_p^{1/1-\gamma}$. This result is shown in the inset of Figure 3. These cut-off effects are only visible for high connectivity levels, though. The network is perfectly scale-free, but only from a local perspective, and repeats itself throughout the entire network for typical scales above z_p [25]. Figure 3 also shows that the obtained *level 0* instantaneous network is well represented by a stretched exponential degree distribution [26] or a power-law with a exponential cut-off for high connectivities [25,27,28]. The network shows a strong dependence on the amount of links at *level 1*. Depending on the strength of the constraints, different cut-offs emerge from the growing process and the power-law obtained with the Barabási-Albert model is recovered when z_c moves the constraint network to a fully-connected graph. Moreover, the topology of *level 1* may influence the topology of *level 0* in different ways, which may not change the average potential degree. For instance, suppose each type to be a country or city which some, as a result of social and economical reasons, interact with more countries or cities than others [29]. To represent this feature, some degree heterogeneity need to be introduced at *level 1*, being the regular network replaced by other class of topologies [3]. This might produce non-trivial results at *level 0*. The typical cut off could become smoother and allows some instances to receive more links than others. From the beginning, there will be nodes that can potentially interact with many more than others – *some nodes are born to become hubs*. Such effects will be tackled in a future work.

4   Conclusions

This last example, together with the even more simplified sexual one, sufficiently illustrates the effects of an extra level of complexity on the growing process of a network, justifying the need for this enlarged perspective. These conceptual additions may contribute to a better understanding of phenomena associated with the structural evolution of complex networks. To the dynamics of the nodes and the evolution of the topology, external constraints, coming from intrinsic features of the nodes and the way they do interact, need to be added. Moreover,

the interplay between the two levels of representation and between the instantaneous and the potential networks remains an open problem in various scientific domains. We have tried in this paper to propose an wide perspective, allowing a better introduction of additional effects which are responsable for shaping networks' evolution. Besides analysing all interacting agents as a whole, different levels of description need to be accounted and their influence investigated, both in the experimental data and in simulation. Furthermore, instead of choosing one unique level of complexity and representation for each problem, a more extended framework has been demonstrated to produce important results that are necessary to understand the intricate network of influences that underlie each complex phenomenon.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47–97 (2002)
2. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science 286(5439), 509–512 (1999)
3. Amaral, L.A., Scala, A., Barthelemy, M., Stanley, H.E.: Classes of small-world networks. Proc. Natl. Acad. Sci. U S A 97(21), 11149–11152 (2000)
4. Tanaka, R., Yi, T.-M., Doyle, J.: Some protein interaction data do not exhibit power-law statistics. FEBS letters 579, 5140–5144 (2005)
5. Dorogotsev, S.N., Mendes, J.F.F.: Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, Oxford (2003)
6. Barabási, L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. Nature Reviews Genetics 5, 101–113 (2004)
7. May, R.M., Lloyd, A.L.: Infection dynamics on scale-free networks. Phys. Rev. E 64(6 Pt 2), 66112 (2001)
8. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Physical Review Letters 86(3200) (2001)
9. Guimerá, R., Amaral, L.A.N.: Functional cartography of complex metabolic networks. Nature 433(895) (2005)
10. Santos, F.C., Pacheco, J.M.: Scale-free networks provide a unifying framework for the emergence of cooperation. Phys. Rev. Lett. 95(9), 098104 (2005)
11. Dorogovtsev, S.N., Mendes, J.F., Samukhin, A.N.: Size-dependent degree distribution of a scale-free growing network. Phys. Rev. E 63(6 Pt 1), 062101 (2001)
12. Vazquez, F., Flamimi, A., Maritan, A., Vespignani, A.: Modeling of protein interaction networks. ComplexUs 1, 38–44 (2003)
13. Wagner, A.: How the global structure of protein interaction networks evolve. Proc. R. Soc. Lond. B. 270, 457–466 (2003)
14. Bersini, H., Lenaerts, T., Santos, F.C.: Growing biological networks: Beyond the gene-duplication model. Journal of Theoretical Biology 241(3), 488–505 (2006)

15. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M.: A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. Nature 403(6770), 623–627 (2000), `http://dx.doi.org/10.1038/35001009`
16. Uetz, P., Finley, J.: From protein networks to biological systems. FEBS letters 579(8), 1821–1827 (2005)
17. Liljeros, F., Edling, C.R., Amaral, L.A., Stanley, H.E., Aberg, Y.: The web of human sexual contacts. Nature 411(6840), 907 (2001)
18. Liljeros, F., Edling, C.R., Amaral, L.A.N.: Sexual networks: implications for the transmission of sexually transmitted infections. Microbes Infect 5(2), 189–196 (2003)
19. Varela, F., Coutinho, A.: Second generation immune network. Immunology Today 12(5), 159–166 (1991)
20. Detours, V., Bersini, H., Stewart, J.: Development of an idiotypic network in shape space. Journal of Theor. Biol. 170, 401–404 (1994)
21. Perelson, A.S.: Immune network theory. Immunological Reviews 10, 5–36 (1989)
22. Perelson, A., Weisbuch, G.: Immunology for physicists. Rev. Mod. Phys. 69, 1219–1268 (1997)
23. Barabási, A.L., Albert, R., Jeong, H.: Mean-field theory for scale-free random networks. Physica A 272, 173–187 (1999)
24. Watts, D.J., Stogatz, S.H.: Collective dynamics of small-world'networks. Nature 393(440) (1998)
25. Rozenfeld, A.F., Cohen, R., Ben-Avraham, D., Havlin, S.: Scale-free networks on lattices. Phys. Rev. Lett. 89(21), 218701 (2002)
26. Xulvi-Brunet, R., Sokolov, I.M.: Evolving networks with disadvantaged long-range connections. Phys. Rev. E 66(2), 026118 (2002)
27. Gastner, M.T., Newman, M.E.J.: The spatial structure of networks. Eur. Phys. J. B 49, 247–252 (2006)
28. Yang, K., Huang, L., Yang, L.: Lattice scale-free networks with weighted linking. Phys. Rev. E 70(1), 015102 (2004)
29. Colizza, V., Barrat, A., Barthelemy, M., Vespignani, A.: Prediction and predictability of global epidemics: the role of the airline transportation network. Proc. Natl. Acad. Sci. U S A 103, 2015–2020 (2006)