*Journal of Complex Networks* (2017) **5**, 694–711
doi: 10.1093/comnet/cnx002
Advance Access Publication on 22 March 2017

# Structural bounds on the dyadic effect

MATTEO CINELLI[†]
*Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico, 1 - 00133 Rome, Italy*
[†]Corresponding author. Email: matteo.cinelli@uniroma2.it

GIOVANNA FERRARO
*Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico, 1 - 00133 Rome, Italy*

AND

ANTONIO IOVANELLA
*Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico, 1 - 00133 Rome, Italy*

Edited by: Piet Van Mieghem

[Received on 5 October 2016; editorial decision on 9 February 2017; accepted on 14 February 2017]

The dyadic effect is a phenomenon that occurs when the number of links between nodes sharing a common feature is larger than expected if the features are distributed randomly on the network. In this article, we consider the case when nodes are distinguished by a binary characteristic. Under these circumstances, two independent parameters, namely dyadicity and heterophilicity are able to detect the presence of the dyadic effect and to measure how much the considered characteristic affects the network topology. The distribution of nodes characteristics can be investigated within a two-dimensional space that represents the feasible region of the dyadic effect, which is bound by two upper bounds on dyadicity and heterophilicity. Using some network structural arguments, we are able to improve such upper bounds and introduce two new lower bounds, providing a reduction of the feasible region of the dyadic effect as well as constraining dyadicity and heterophilicity within a specific range. Some computational experiences show the bounds effectiveness and their usefulness with regards to different classes of networks.

*Keywords*: complex networks; dyadic effect; upper and lower bound.

## 1. Introduction

Complex systems modelled as networks exhibit global structures that are commonly affected by the characteristics of their founding elements. Indeed, the properties of these elements often correlate with the architecture of the observed systems ([2, 26]). This article is devoted to the situations where nodes themselves have peculiar properties that carry significant information regarding their role within the network topology. In literature ([8, 25]), the tendency of nodes to link with others that are similar to themselves is a phenomenon called *homophily*, which affects the dyadic similarities between nodes and creates correlated patterns among neighbours. The nodes tendency to connect with each other also relates to the concept of assortative mixing [27] which describes correlations because of some nodes properties.

Park and Barabási [30] noted that, when nodes in a network fit within two distinct groups according to their characteristics, two different parameters, namely *dyadicity* and *heterophilicity* are required to identify the relations between the network topology and nodes features.

The dyadic effect has been considered in order to assess the functional role of nodes within biological networks such as, for instance, in gene-gene interaction in statistical epistasis networks [20], in phenome-genome networks [21] in disease-phenotype networks [22] and in protein-protein interaction networks ([10, 34]) where numerous characteristics are studied to evaluate genetic interactions. Nodes' characteristics are investigated also in inter-organizational innovation networks ([14–16]) where partnerships agreements of technological transfer among countries are related to innovation indices.

The investigation of mixing patterns within networks can be performed using at least two different approaches. The first is based on the assortativity coefficient [7, 13, 24, 32, 35] that represents a *second order network metric* [29], able to quantify through a unique index the presence of homophilic patterns in the whole network for either scalar (e.g. degree) or discrete node characteristics (e.g. gender). The second approach constrained to cases when each node is characterized by a binary feature [30] goes towards a microscopic level of detail being based on the study of dyads. The analysis of the interplay among different dyads is able to unveil the node-network correlation through the use of two coefficients that separately quantify the level of homogeneity and heterogeneity in networks. This lets room to the identification of local aspects for instance, we can observe a dyadic and heterophilic network being globally disassortative as we will show later on. Indeed the relationship between assortative mixing and dyadic effect requires further investigation.

The methodology presented in [30] is able to study exhaustively all the configurations of a binary characteristic on a network using a phase diagram, which lies into a two dimensional space constrained by certain network related bounds. However, current bounds are computed a-priori considering either particular networks arguments or the number of featured nodes, which results in a space that is often much larger than necessary. Another important issue is related to the phase diagram computational complexity which grows exponentially in the number of nodes, implying some limitations in real applications. The literature has attempted to overcome these difficulties by using heuristics or statistical methods. For instance, in [30] it is reported a heuristic method able to identify extremal configurations; in [20–22], and [34] statistical methods are used to infer the existence of some configurations in computational biology, while in [3] entropy-based measures are used to globally assess the relevance of nodes characteristics.

The contribution of this article lies in the improvement of current upper bounds through considerations related to structural arguments surrounding a given network. The reduction of the two dimensional space is performed by not only considering the two upper bounds but by also introducing two lower bounds. We also present the analytical reasonings and we test their behaviour on different classes of networks.

New bounds foundations are rooted in the degree sequence which can be easily extracted from any network. Although the results we obtain in the space reduction can be valuable, we make use of the straightforward relationships that allow us to compute such bounds independently of the network size.

We provide a reduction of the dyadic effect feasible region, which can be used in all the applications where the degree of correlation of the nodes characteristics with the network topology is evaluated considering empirical arguments instead of computing the phase diagram such as, for instance, in [20] and [4]. Moreover, by introducing the four bounds we are limiting the values of dyadicity and heterophilicity to lie in a range that can be easily computed.

The article is organized as follows: Section 2 gives the problem settings; Section 3 shows the upper and lower bounds; Section 4 contains the computational analysis; Section 5 presents the conclusions.

## 2. Problem settings

### 2.1 *Theoretical background*

The classical mathematical abstraction of a network is a graph $G$. Let $G = (V, E)$ be a graph composed of a set $V$ of $N$ nodes and a set $E$ of $M$ edges that defines the relationship among these nodes. Herein $G$ is considered undirected, unweighted, connected and simple, i.e. loops and multiple edges are not allowed.

The degree $d_i$ of a node $i \in V$ is defined as the number of edges in $E$ incident to $i$. The nodes degrees listed in an non-increasing order are referred to as *degree sequence* $D_G$ and, as we recall, for every connected graph holds the Degree-Sum Formula or *Handshaking Lemma*, $\sum_{i=1}^{N} d_i = 2M$. A *graphic sequence* is defined as a list of non-negative numbers which is the degree sequence of certain simple graphs. A graph $G$ with degree sequence $D_G$ is called a *realization* of $D_G$.

A generic list $L$ of non-negative numbers is not necessarily a graphic sequence. Indeed, a necessary and sufficient condition is that $\sum_{i=1}^{N} d_i$ is even and $\sum_{i=1}^{N} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min\{k, d_i\}$, $1 \leq k \leq N$ [11]. The problem of discovering if $L$ is a graphic sequence is called *Graph Realization Problem* ([18, 19]).

Given an integer $n \leq N$, we consider the subsequence of the first $n$ elements of $D_G$ calling it $D_G^H(n) \subseteq D_G$ as the *head* of $D_G$, and the subsequence of the last $n$ elements of $D_G$ calling it $D_G^T(n) \subseteq D_G$ as the *tail* of $D_G$.

A *clique* $K_n$ is a complete subgraph of $G$ of dimension $n$, i.e. a subgraph of $n$ mutual interconnected nodes. The problem of finding the clique of highest cardinality in $G$ is a well-known *NP*-problem [17]. Since the degree of nodes in a clique of cardinality $n$ is at least $n-1$, a necessary condition for the existence of such clique is that $D_G$ contains at least $n$ nodes of degree $d_i \geq n-1$, otherwise the graph $G$ cannot contain such clique.

A *star* $S_n$ is a subgraph of $G$ of dimension $n$ showing one node with degree $n-1$ and the others $n-1$ having degree 1.

Traditionally, in graph theory, edges have two endpoints since, by definition, they represent a reciprocal relationship between two nodes. In the literature of graph realization problem, as well as that of dynamic graphs, it is possible to admit half-edges anchored at one node of a degree sequence while the other endpoint is free. This particular object is called *stub* [28]; when two stubs of two distinct nodes connect, then a classical edge is realized [23].

### 2.2 *Nodes characteristics and dyadic effect*

Herein, we refer to a given characteristic $c_i$, which can assume the values 0 or 1, for each $i \in N$. Consequently, $N$ can be divided into two subsets: the set of $n_1$ nodes with characteristic $c_i = 1$, the set of $n_0$ nodes with characteristic $c_i = 0$; thus, $N = n_1 + n_0$. We distinguish three kinds of *dyads*, i.e. edges and their two end nodes, in the network: (1 − ï¿½1), (1 − ï¿½0) and (0 − ï¿½0) as depicted in the Fig. 1.

We label the number of each dyad in the graph as $m_{11}$, $m_{10}$, $m_{00}$, respectively. Hence, $M = m_{11} + m_{10} + m_{00}$. We consider $m_{11}$ and $m_{10}$ as independent parameters that represent the dyads containing nodes with characteristic 1.

Let $D_G$ be the degree sequence of $G$, we can use $n_1$ and $n_0$ to define its heads $D_G^H(n_1)$ or $D_G^H(n_0)$ and the tails $D_G^T(n_1)$ or $D_G^T(n_0)$ such as $D_G = D_G^H(n_1) \cup D_G^T(n_0)$ or $D_G = D_G^H(n_0) \cup D_G^T(n_1)$. These partitions

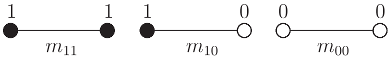

FIG. 1. Types of dyads.

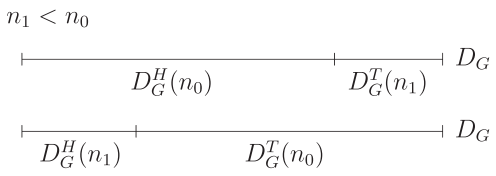

FIG. 2. Two different partitions when $n_1 < n_0$.

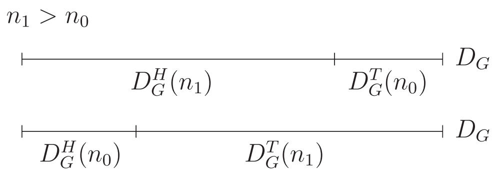

FIG. 3. Two different partitions when $n_1 > n_0$.

of the degree sequence are given arbitrarily assigning the characteristic $c_i = 1$ to the $n_1$ nodes with the highest degree or to the $n_1$ nodes with lowest degree or vice versa. Such partitions are reported in Figs 2 and 3, distinguishing the case in which $n_1 < n_0$ or $n_1 > n_0$. We make this construction in order to use it in Section 3.

When nodes in a network fit within two distinct groups according to their characteristics, two different parameters are required to determine the existence of the relations between the network topology and the nodes features [30]. In many systems, the number of edges between nodes sharing a common characteristic is larger than expected if the characteristics are distributed randomly on the graph; this phenomenon is called the *dyadic effect* [33]. If a casual setting among the $N$ nodes is considered, where any node has an equal chance of having the characteristic 1, the expected values of $m_{11}$ and $m_{10}$ are [30]:

$$\overline{m}_{11} = \binom{n_1}{2}\delta = \frac{n_1(n_1-1)}{2}\delta \tag{2.1}$$

$$\overline{m}_{10} = \binom{n_1}{1}\binom{n_0}{1}\delta = n_1(N-n_1)\delta, \tag{2.2}$$

where $\delta$ is the density and is equal to $\delta = 2M/N(N-1)$. The relevant deviations of $m_{11}$ and $m_{10}$ from the expected values $\overline{m}_{11}$ and $\overline{m}_{10}$ denote that the characteristic 1 is not randomly distributed ([8, 30]). Such deviations can be calculated through the ratios of dyadicity $D$ and heterophilicity $H$ defined as:

$$D = \frac{m_{11}}{\overline{m}_{11}} \tag{2.3}$$

$$H = \frac{m_{10}}{\overline{m}_{10}} \tag{2.4}$$

If the characteristic is dyadic, $D > 1$, it means that nodes with the same characteristics tend to link more tightly among themselves than expected in a random configuration. Conversely when $D < 1$, the characteristic is anti-dyadic, indicating that similar nodes tend to connect less densely among themselves than expected in a random configuration. The characteristic is defined as heterophilic, with a value $H > 1$, highlighting that nodes with the same features have more connections to nodes with different

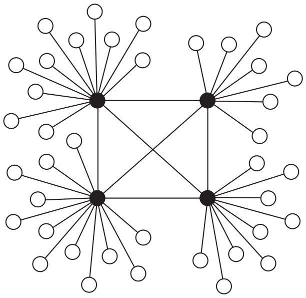

FIG. 4. Example

TABLE 1 *Example measures*

| | |
|---|---|
| $N = 43$ | $M = 45$ |
| $\delta = 0.05$ | $n_1 = 4$ |
| $\overline{m}_{11} = 0.3$ | $\overline{m}_{10} = 8.16$ |
| $r_{discrete} = -0.76$ | $r_{scalar} = -0.75$ |
| $D = 20$ | $D_{max} = 20$ |
| $H = 4.8$ | $H_{max} = 19.11$ |

characteristics than expected randomly. On the contrary, with a value $H < 1$, the characteristic is defined as heterophobic, meaning that nodes with certain characteristics have fewer links to nodes with different characteristics than expected randomly.

In [30], it is established that $m_{11}$ and $m_{10}$ cannot assume arbitrary values, as there are indirect constraints due to the network structure. Indeed, $m_{11}$ cannot exceed

$$UBm_{11} = min(M, \binom{n_1}{2}) \tag{2.5}$$

and $m_{10}$ cannot be larger than

$$UBm_{10} = min(M, n_1 n_0) \tag{2.6}$$

where *UB* stands for upper bound.

Lastly, we provide an example of the aforementioned measures being applied in order to shed more light on the differences between the assortativity coefficient $r$ and the metrics $D$ and $H$. We take into account a network with $N = 43$, $M = 45$ and $n_1 = 4$, where we have the four higher degree nodes having characteristic $c_i = 1$ (see Fig. 4). As reported in Table 1, the network displays a strong disassortative mixing either to degree $r_{scalar} = -0.75$ or to the discrete characteristic $r_{discrete} = -0.76$ (values are computed using the formulas as in [27]) meaning that similar nodes tend to avoid each other. The disassortative mixing at global level hides the presence of an important local substructure (the so called rich-club [5, 36]) in which similar nodes are tightly connected. Note that the value of dyadicity is $D = D_{max} = 20$, which is the maximum achievable value, while the disassortative mixing is confirmed by the value of heterophilicity $H = 4.8$.

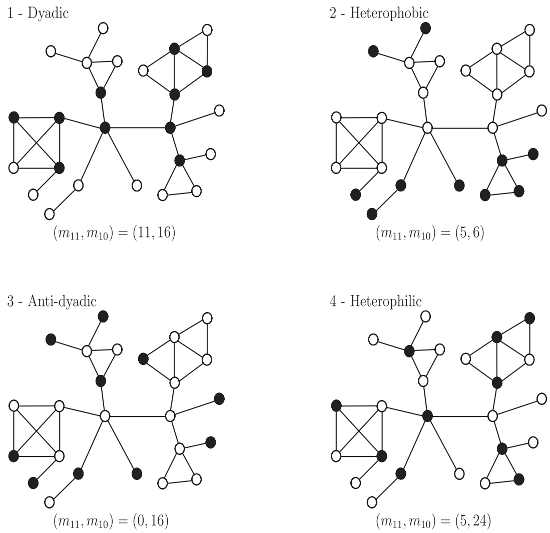
1 - Dyadic
$(m_{11}, m_{10}) = (11, 16)$
2 - Heterophobic
$(m_{11}, m_{10}) = (5, 6)$
3 - Anti-dyadic
$(m_{11}, m_{10}) = (0, 16)$
4 - Heterophilic
$(m_{11}, m_{10}) = (5, 24)$

FIG. 5. Configurations of four extreme points on the phase diagram.

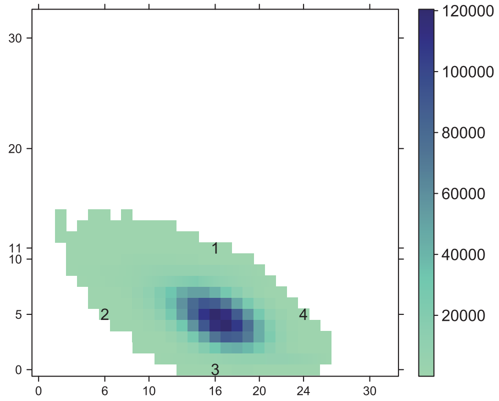

FIG. 6. The phase diagram of possible values of ($m_{11}$, $m_{10}$ when $n_1 = 10$). The values of $m_{11}$ and $m_{10}$ referred to the four configurations of Fig. 5 are reported.

## 2.3 *The phase diagram*

One instrument to investigate the correlation among the distribution of a given property $c$ and the underlying network structure is the phase diagram which, in general, describes the admissible configurations in the graph.

We consider, as an example, the graph shown in Fig. 5 that depicts a network with 25 nodes and 32 edges of which $n_1 = 10$ black nodes are randomly distributed.

The corresponding phase diagram in Fig. 6 describes the distribution of a random feature in the system. It should be noted that the reported phase diagram shows only a subarea.

The phase diagram presents all the admissible combinations of $m_{10}$ ($x$-coordinate) and $m_{11}$ ($y$-coordinate) and each corresponding square collects the number of assignments of $n_1$ nodes over the set $N$ for every fixed $m_{10}$ and $m_{11}$. There is a direct correspondence among the $m_{10}$ and $m_{11}$ axis and, respectively, $H$ and $D$, since the values are related through means of Equations 2.3 and 2.4. Moreover, $m_{11}$ from 0 to $UBm_{11}$ and $m_{10}$ ranges from 0 to $UBm_{10}$. Correspondingly, $D$ ranges from 0 to $D_{max} = UBm_{11}/\overline{m}_{11}$ and $H$ ranges from 0 to $H_{max} = UBm_{10}/\overline{m}_{10}$. For a given $m_{11}$ and $m_{10}$, each square has a darkness proportional to the degeneracy of the configuration and an open square means that is not possible to place $n_1$ nodes consistently with the fixed values and constraints imposed by the network topology.

Beside such squares, the phase diagram has some other meaningful areas to discuss. In particular, the high degeneracy squares are considered as the most typical configurations for a random distribution of a property $D = H = 1$; and the phase boundaries squares map atypical configurations. For such phase boundaries different layouts are recognizable. Indeed, in Fig. 5, four possible configurations are represented (where the point of each configuration is correspondingly numbered in the phase diagram of Fig. 6): $D \gg 1$ is a dyadic case where black nodes concentrate in a central cluster of the graph which maximizes $m_{11}$; $D \ll 1$ is an anti-dyadic configuration where black nodes tend to be farther apart; $H \ll 1$ is an heterophobic configuration where black nodes are located in the peripheral area which minimize $m_{10}$; and, $H \gg 1$ is an heterophilic configuration where black nodes correspond to the most connected nodes so that the edges with white nodes are maximized.

The graphical nature of the phase diagram allows for easier observation of the distribution of the nodes characteristics, however, as the number of the possible configurations increases exponentially with $N$, the phase diagram is hard to compute for large networks. Moreover, since $n_1 \in [0, N]$ it can change while the network structure remains the same. Indeed the number of nodes showing a certain characteristic can change, or different characteristics can be studied by varying $n_1$, such as in [20]. Therefore, a complete analysis may require a sequence of $N + 1$ phase diagrams computed for each value of $n_1$ as in [4]. An outcome of the latter case for a graph with 25 nodes and 32 edges is shown in Fig. 7.

## 3. Upper and lower bounds

In this section, we propose an extension of upper bounds (2.5) and (2.6) for a graph $G$ given $N$, $M$ and $n_1 \in [0, N]$. Moreover, we propose two lower bounds (*LB*) to $m_{11}$ and $m_{10}$. In other words, we want to restrict the feasible region of the dyadic effect for a graph $G$ as much as possible, excluding non-admissible configurations.

We can notice that in [30] no lower bounds are provided, thus both bounds on $m_{11}$ and $m_{10}$ are assumed to be zero at minimum.

### 3.1 *Upper bound $UBm_{11}$*

Equation (2.5) states that the maximum number of $m_{11}$ within a network is equal to the minimum between two quantities: the number of network edges (meaning that all the edges are $m_{11}$ and $n_1 = N$); the number of edges within a clique $K_{n_1}$, i.e. a complete subgraph with $n_1$ nodes within $G$.

The rationale behind the latter statement is that the upper bound is pushed to the maximum value when $G$ is supposed to contain $n_1$ nodes arranged in a clique.

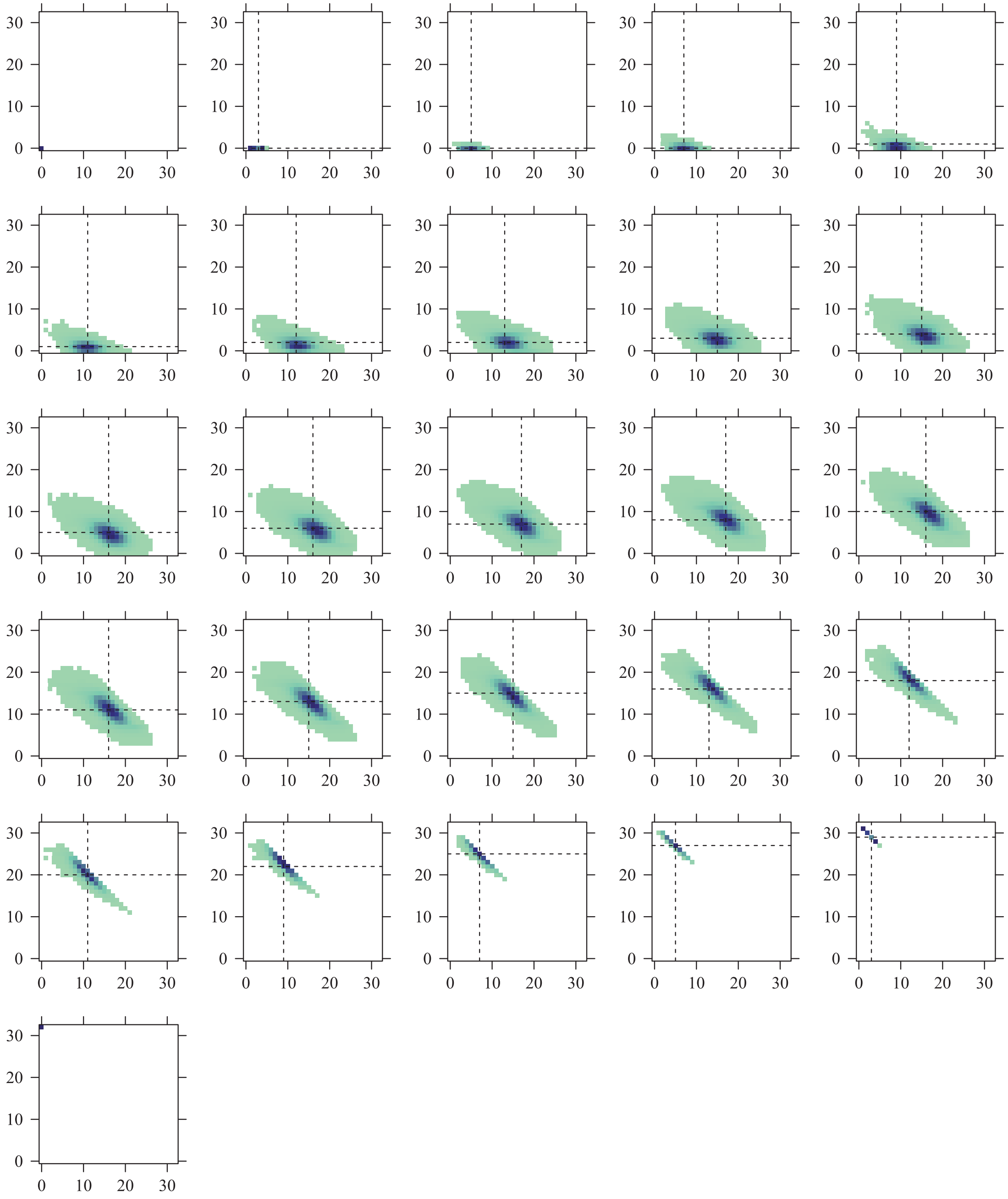

FIG. 7. The sequence of the phase diagrams of the network depicted in Fig. 5 when $n_1$ grows from 0 to $N$. The value of $n_1$ increases starting from left to right and from the upper to the lower side of the figure therefore it should be read row by row and dotted lines indicate the coordinates of $D = H = 1$. The sequence shows the various shapes assumed by the phase diagram, thus its sensitivity to changes of $n_1$ in terms of covered area.

PROPOSITION 3.1 Let us consider the degree sequence $D_G$ of the graph $G$ with $n_1$ nodes having the characteristic equal to 1. The upper bound $UBm_{11}$ on the number of edges $m_{11}$ is:

$$UBm_{11} = min\left(M, \binom{n_1}{2}, \left\lceil \sum_{i \in D_G^H(n_1)} \frac{\min(d_i, n_1 - 1)}{2} \right\rceil\right) \tag{3.1}$$

*Proof.* Considering the degree sequence, we distinguish two different cases based on the fact that $D_G$ may or may not contain $n_1$ nodes of degree at least $n_1 - 1$.

1. *$D_G$ contains $n_1$ nodes of degree at least $n_1 - 1$.*

   In this case, the necessary condition for the existence of $K_{n_1}$ holds and we can suppose, as a worst case, the realization of a clique $K_{n_1}$ considering $D_G^H(n_1)$. Thus, the bound given in formula (2.5) is the tightest.

   Note that if in $D_G$ it is possible to realize a clique $K_{n_1}$, its nodes can be considered the same as those in $D_G^H(n_1)$. Indeed, $D_G$ is an ordered sequence and if $D_G^H(n_1)$ does not realize $K_{n_1}$, none other subsequence into $D_G$ can realize it. Therefore, we refer our analysis to the subsequence $D_G^H(n_1)$.

2. *$D_G$ does not contain $n_1$ nodes of degree at least $n_1 - 1$.*

   As $K_{n_1}$ is the densest possible realization in $D_G^H(n_1)$, we can similarly search for the densest possible realization actually feasible in the subsequence when the clique is not realizable. By construction, $D_G$ is graphic but a subsequence cannot be graphic albeit the sum of its elements is even. Indeed, $D_G^H(n_1)$ is a graphic sequence only if conditions reported in [11, 18, 19] hold; otherwise, it is not graphic.

   (a) *$D_G^H(n_1)$ is graphic.*

   In this situation, the densest hypothetical realization of $D_G^H(n_1)$ is a graph in which the handshaking lemma holds, thus the number of its edges is $m_{11} = \sum_{i \in D_G^H(n_1)} \min(d_i, n_1 - 1)/2$ because each one of the $n_1$ nodes has its degree bounded by the value $n_1 - 1$.

   (b) *$D_G^H(n_1)$ is not graphic.*

   Since we are searching for upper bounds, the situation in which $D_G^H(n_1)$ is not graphic can be managed through an overestimation of the handshaking lemma. In this case, we consider the densest hypothetical realization of a simple graph that involves the maximum number of stubs corresponding to $\min(d_i, n_1 - 1)$ for any $i \in D_G^H(n_1)$. In order to obtain the $UB$ value, we consider the involvement of all the elements in $D_G^H(n_1)$, ceiling the sum if odd. Through this procedure the handshaking lemma holds and we can compute the number of edges as $m_{11} = \lceil \sum_{i \in D_G^H(n_1)} \min(d_i, n_1 - 1)/2 \rceil$.

   Summarizing all the considerations thus far, $UBm_{11}$ can be written as in Formula (3.1).

□

### 3.2 *Upper bound $UBm_{10}$*

Equation (2.6) states that the maximum number of $m_{10}$ is equal to the minimum between $M$, meaning that all the edges are $m_{10}$ and thus there are no adjacent $n_1$, and the number of edges within a set of $n_1$ stars of degree $n_0$ (or $n_0$ stars of degree $n_1$).

In more detail, the second element in the upper bound formula implies that all the $n_1$ nodes are arranged in order to be the central nodes of a set of stars $S_{n_0+1}$ with non-adjacent central vertices and of degree $n_0$, or vice versa.

Proposition 3.2 Let us consider the degree sequence $D_G$ of the graph $G$ with $n_1$ nodes having the characteristic equal to 1. The upper bound $UBm_{10}$ on the number of edges $m_{10}$ is:

$$UBm_{10} = min\left(M, n_1 n_0, min\left(\sum_{i\in D_G^H(n_1)} \min(d_i, n_0), \sum_{i\in D_G^H(n_0)} \min(d_i, n_1)\right)\right) \tag{3.2}$$

*Proof.* If the graph $G$ can contain $n_1$ stars $S_{n_0+1}$, i.e. in $D_G$ are present at least $n_1$ elements with $d_i \geq n_0$ then the maximum number of stars is theoretically allowed and the bound given in [30] can be considered the tightest. Otherwise if such stars do not exist we can take into account the set of stars that is actually realizable using the degree sequence of $G$.

Clearly, the same reasoning can be applied when considering $n_0$ instead of $n_1$. For any fixed $n_1$, stars can be realized with all the central nodes having the characteristic $c_i = 1$ and the other elements having $c_i = 0$ and vice versa.

When maximizing $m_{10}$, we ask for the set of stars with non-adjacent central nodes that brings $m_{11}$ to be the minimum, i.e. equal to 0. Under these considerations, we are faced with three different situations: $n_1 < n_0$, $n_1 > n_0$ and $n_1 = n_0$.

1. *Suppose that $n_1 < n_0$.*

   In this case, we can partition the degree sequence of $G$ as $D_G = D_G^H(n_1) \cup D_G^T(n_0)$ or as $D_G = D_G^H(n_0) \cup D_G^T(n_1)$ (see Fig. 2).

   When $D_G = D_G^H(n_1) \cup D_G^T(n_0)$, the elements in $D_G^H(n_1)$ show a number of stubs equal to the sum of their degree that, in order to realize edges $m_{10}$, have to find their endpoints in $D_G^T(n_0)$. Three cases are admissible:

   - $\sum_{i\in D_G^H(n_1)} d_i = \sum_{i\in D_G^T(n_0)} d_i$: all stubs in $D_G^H(n_1)$ have an endpoint in $D_G^T(n_0)$. In this case the realization on $D_G$ has $M = m_{10} = \sum_{i\in D_G^H(n_1)} d_i$.
   - $\sum_{i\in D_G^H(n_1)} d_i < \sum_{i\in D_G^T(n_0)} d_i$: all stubs in $D_G^H(n_1)$ have an endpoint in $D_G^T(n_0)$ but some stubs in $D_G^T(n_0)$ remain free. In this case, the realization on $D_G$ has $m_{10} = \sum_{i\in D_G^H(n_1)} d_i$, while $M$ contains some $m_{00}$.
   - $\sum_{i\in D_G^H(n_1)} d_i > \sum_{i\in D_G^T(n_0)} d_i$: not all stubs in $D_G^H(n_1)$ have an endpoint in $D_G^T(n_0)$. In this case, the realization $D_G$ has all stubs in $D_G^T(n_0)$ saturated while some residual stubs in $D_G^H(n_1)$ can create edges between those nodes not involved in stars. In this case, $m_{11}$ may be different to zero and an overestimation is $m_{10} = \sum_{i\in D_G^H(n_1)} d_i$.

When $D_G = D_G^H(n_0) \cup D_G^T(n_1)$ three cases can be discussed:

- $\sum_{i \in D_G^H(n_0)} d_i = \sum_{i \in D_G^T(n_1)} d_i$: since $n_0$ is greater than $n_1$ and $D_G$ is in non-increasing order, this case is not admissible.
- $\sum_{i \in D_G^H(n_0)} d_i < \sum_{i \in D_G^T(n_1)} d_i$: again, this case is not admissible for the same reason as above.
- $\sum_{i \in D_G^H(n_0)} d_i > \sum_{i \in D_G^T(n_1)} d_i$: not all stubs in $D_G^H(n_0)$ have an endpoint in $D_G^T(n_1)$. In this case, the realization on $D_G$ has all stubs in $D_G^T(n_1)$ saturated while some residual stubs in $D_G^H(n_0)$ can create edges that increase the number of $m_{00}$. Thus, an overestimation is $m_{10} = \sum_{i \in D_G^H(n_0)} d_i$.

Summarizing, we provided certain overestimations on the number of $m_{10}$ and the minimum among them is our upper bound when $n_1 < n_0$.

2. *Suppose that* $n_1 > n_0$.

In this case, we can partition the degree sequence of $G$ as $D_G = D_G^H(n_0) \cup D_G^T(n_1)$ or as $D_G = D_G^H(n_1) \cup D_G^T(n_0)$ (see Fig. 3) and all the same considerations of above can be repeated, using caution to invert $n_0$ and $n_1$. Again, we obtain certain overestimations on the number of $m_{10}$ and the minimum among them is our upper bound.

3. *Suppose that* $n_1 = n_0$

Then in this case $D_G = D_G^H(n_1) \cup D_G^T(n_0) = D_G^H(n_0) \cup D_G^T(n_1)$. Considering the first partition, three situation can be discussed:

- $\sum_{i \in D_G^H(n_1)} d_i = \sum_{i \in D_G^T(n_0)} d_i$: this happens only when $G$ is regular.
- $\sum_{i \in D_G^H(n_1)} d_i < \sum_{i \in D_G^T(n_0)} d_i$: this case is not admissible.
- $\sum_{i \in D_G^H(n_1)} d_i > \sum_{i \in D_G^T(n_0)} d_i$: not all stubs in $D_G^H(n_1)$ have an endpoint in $D_G^T(n_0)$. In this case, the realization on $D_G$ has all stubs in $D_G^T(n_0)$ saturated while some residual stubs in $D_G^H(n_1)$ can create edges between nodes not involved in stars. In this case, $m_{11}$ can be different to zero and an overestimation is $m_{10} = \sum_{i \in D_G^H(n_1)} d_i$.

Such considerations can be repeated for the second partition of $D_G$.

Moreover, knowing the size of the two partitions, we can further bound the introduced quantities and, consequently, $m_{10}$. Indeed, in order to realize $m_{10}$, each element in $D_G^H(n_1)$ or $D_G^T(n_1)$ can be connected at most to $n_0$ others while each element in $D_G^H(n_0)$ or $D_G^T(n_0)$ can be connected at most to $n_1$ others. Thus every $d_i \geq n_0$ in $D_G^H(n_1)$ or $D_G^T(n_1)$ is actually bounded by $n_0$ while every $d_i \geq n_1$ in $D_G^H(n_0)$ or $D_G^T(n_0)$ is actually bounded by $n_1$; furthermore the residual degree of each $d_i$ does not contribute to the formation of $m_{10}$.

Finally, the value for the upper bound on the number of $m_{10}$ can be written as in formula (3.2). □

### 3.3 *Lower bound LBm*$_{11}$

We propose a lower bound of $m_{11}$, observing under which conditions a hypothetical graph realization of $D_G$ exists that contains at least some $m_{11}$. Such quantity is considered as an underestimation of $m_{11}$ in the original $G$.

Proposition 3.3 Let us consider the degree sequence $D_G$ of the graph $G$ with $n_1$ nodes having the characteristic equal to 1. The lower bound $LBm_{11}$ on the number of edges $m_{11}$ is:

$$LBm_{11} = \max\left(0, \left\lfloor \frac{\sum_{i\in D^T_G(n_1)} d_i - \sum_{i\in D^H_G(n_0)} d_i}{2} \right\rfloor\right) \tag{3.3}$$

*Proof.* Given $n_1$, let consider again the two possible partitions of the degree sequence $D_G = D^H_G(n_1) \cup D^T_G(n_0) = D^H_G(n_0) \cup D^T_G(n_1)$. The following three cases hold for $n_1 \gtreqless n_0$:

- $\sum_{i\in D^H_G(n_1)} d_i = \sum_{i\in D^T_G(n_0)} d_i$ or $\sum_{i\in D^H_G(n_0)} d_i = \sum_{i\in D^T_G(n_1)} d_i$: when admissible, stubs have endpoints in different partitions and $M = m_{10}$; thus, $m_{11} = 0$.
- $\sum_{i\in D^H_G(n_1)} d_i < \sum_{i\in D^T_G(n_0)} d_i$ or $\sum_{i\in D^H_G(n_0)} d_i > \sum_{i\in D^T_G(n_1)} d_i$: when admissible, there is no room for residual degree in the partition of nodes with the characteristic $c_i = 1$; thus, $m_{11} = 0$.
- $\sum_{i\in D^H_G(n_1)} d_i > \sum_{i\in D^T_G(n_0)} d_i$ or $\sum_{i\in D^H_G(n_0)} d_i < \sum_{i\in D^T_G(n_1)} d_i$: when admissible, some stubs in the partition of nodes with the characteristic $c_i = 1$ can link among themselves.

The third case can happen when $\sum_{i\in D^H_G(n_1)} d_i - \sum_{i\in D^T_G(n_0)} d_i > 0$ or $\sum_{i\in D^T_G(n_1)} d_i - \sum_{i\in D^H_G(n_0)} d_i > 0$. Note that, since we are searching for the minimum number of stubs able to make $m_{11} \neq 0$ and that for any given $n_1$ $\sum_{i\in D^H_G(n_1)} d_i \geq \sum_{i\in D^T_G(n_1)} d_i$, we can restrict our analysis to the cases when $D_G = D^H_G(n_0) \cup D^T_G(n_1)$. Therefore, a lower bound on the number of $m_{11}$ is given by selecting the maximum value between 0 and an underestimation of the possible edges created in the partition $D^T_G(n_1)$, such as in formula (3.3).

□

### 3.4 *Lower bound $LBm_{10}$*

When $n_1 = 0$ or $n_1 = N$, the number of $m_{10}$ is trivially 0. In all other cases, any possible connected realization of $D_G$ contains at least an edge with endpoints with different characteristics, i.e. $m_{10} \geq 1$.

In order to discuss a lower bound on $m_{10}$ that, in some cases, overcomes the given inequality from above, we take into account certain arguments based on the realizability of a complete, or at least densest, subgraph from $D_G$, similarly to as in Section 3.1.

Proposition 3.4 Let us consider the degree sequence $D_G$ of the graph $G$ with $n_1$ nodes having the characteristic equal to 1. The lower bound $LBm_{10}$ on the number of edges $m_{10}$ is:

$$LBm_{10} = \begin{cases} 0 & \text{if } n_1 = 0, N \\ max\left(1; \sum_{i\in D^T_G(n_1)} d_i - n_1(n_1 - 1)\right) & \text{if } n_1 \in (0, N) \end{cases} \tag{3.4}$$

*Proof.* When $n_1 \in (0, N)$ we distinguish two cases determined by whether $D_G$ contains $n_1$ nodes of degree at least $n_1 - 1$ that allows for a realization of a clique $K_{n_1}$.

1. *$D_G$ does not contain $n_1$ nodes of degree at least $n_1 - 1$.*

   In this case, we can suppose that the densest hypothetical realization of a simple graph involving the maximum number of stubs is the same as in the second case of the proof of Proposition 3.1.

Such realization contains the maximum number of $m_{11}$ and at least one residual stub has to find its endpoint in one of the nodes of the remaining part of $D_G$ because any realization should be connected. Therefore, this case leads us to consider $m_{10} = 1$.

2. *$D_G$ contains $n_1$ nodes of degree at least $n_1 - 1$.*

   In this case it can hypothetically realize $K_{n_1}$. In order to search for a lower bound on $m_{10}$, we take into account the subsequence in $D_G$ which contains the minimum number of stubs, i.e. $D_G^T(n_1)$.

   If a clique can be realized within $D_G^T(n_1)$, then $m_{10} = \sum_{i \in D_G^T(n_1)} d_i - n_1(n_1 - 1)$. Indeed, every stub that constitutes the residual degree will find its endpoint in a node with $c_i = 0$. Otherwise, $\sum_{i \in D_G^T(n_1)} d_i < n_1(n_1 - 1)$, a clique cannot be realized and $m_{10} = 1$. □

### 3.5 *Bounds implications on dyadic effect and its applications*

In Section 2.3, we introduced range values for $D$ and $H$ and their relationships with the corresponding upper bounds. Since we defined two new formulas for $UBm_{11}$ and $UBm_{10}$, $D_{max}$ and $H_{max}$ may assume lower values. Moreover, the introduction of $LBm_{11}$ and $LBm_{10}$ results in the definition of two new quantities, i.e. $D_{min} = LBm_{11}/\overline{m}_{11}$ and $H_{min} = LBm_{10}/\overline{m}_{10}$. Thus, we can state that:

PROPOSITION 3.5 Given a simple graph $G$ with $n_1$ nodes having the characteristic equal to 1, the dyadic effect is bounded as follow: $D_{min} \leq D \leq D_{max}$ and $H_{min} \leq H \leq H_{max}$.

This proposition has a main implication. Indeed, since the dyadic effect has been used to quantify homophily, the proposition sets the bounds on nodes tendency to connect with others similar to themselves by using information on the network itself instead of a priori combinatorial arguments. In fact, the bounds presented in [30] are valid for every graph of $N$ nodes, $M$ edges and a fixed integer $n_1$, while the bounds presented in Section 3 depend on the graphic sequence of the given graph, thus are valid for the set of all graphs having the same $D_G$. Such set is still wide but has a tighter relationship with the graph under observation.

Regarding the applications, when a network and a set of characteristics are given, it is straightforward to compute for each characteristic the point of maximum degeneration $D = H = 1$, the values $D$, $H$ and through the use of the four bounds, $D_{min}$, $D_{max}$, $H_{min}$ and $H_{max}$. This approach can be useful in many applications, such as in [20–22] and [34] where the phase diagram is hard to compute. In such contexts, statistical approaches are used to gather information on the correlation between nodes characteristics and the network topology by looking for the relative distance of the point $(D, H)$ from the point of maximum degeneration.

The introduction of the new bounds makes any comparison within the two dimensional space, defined originally in [30] and improved in the previous section, more reliable as they are performed on measures that are deeply related to the structure of the analysed graph.

## 4. Empirical evidence

Herein, we show empirical evidence computing upper and lower bounds as presented in Section 3 for different networks. In particular, we extensively study the test graph given in Fig. 5 in order to provide an evaluation of the feasible region reductions, then we provide results on different instances in order to observe the behaviour of different bounds.

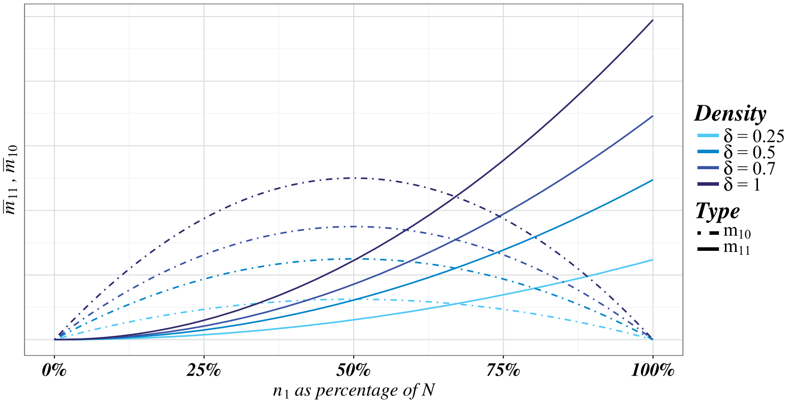

FIG. 8. $\overline{m}_{11}$ and $\overline{m}_{10}$ values as a function of the fraction of $n_1$ nodes on $N$.

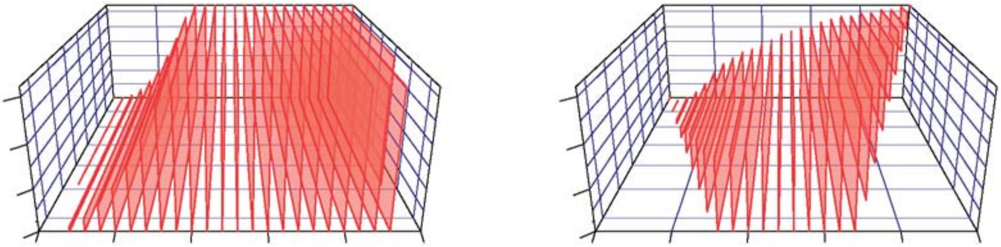

FIG. 9. Areas computed with old and new bounds as a function of $n_1$ nodes on $N$ ($x$ axis). The $y$ and $z$ axes represent $m_{11}$ and $m_{10}$ respectively.

The data processing, network analysis and all simulations were performed using the software *R* [31] with the *igraph* package [6].

All analysis has been conducted considering the given graph $G$ with $N$ nodes and $M$ edges and the value of $n_1$ ranging from 0 to $N$. Values of $\overline{m}_{11}$ and $\overline{m}_{10}$ were computed straightforwardly and independently of $N$ by means of formulas (2.1) and (2.2) and Fig. 8 shows their values as a function of the fraction of $n_1$ nodes on $N$ and for different values of the density $\delta$.

### 4.1 *Analysis of the test graph*

We begin by illustrating the shrinkage of the area in which the phase diagram lies by using the test graph with $N = 25$ and $M = 32$. The areas in Fig. 9 represent the sequence of feasible regions for $n_1$ varying from 0 to $N$, each bounded by the correspondent values of $UBm_{10}$, $LBm_{10}$, $UBm_{11}$ and $LBm_{11}$. Comparing the areas in Fig. 9, we can immediately notice the difference between the feasible regions provided by old and new bounds as well as the consequent improvements mostly appreciable for high and low values of $n_1$. In Fig. 9, we observe how, by applying the bounds, the feasible region changes together with the various shapes of the $N + 1$ phase diagrams as shown in Fig. 7. Moreover, the areas evolve following a trajectory which reflects the trend observed in the curves of the expected values of $m_{11}$ and $m_{10}$ in Fig. 8.

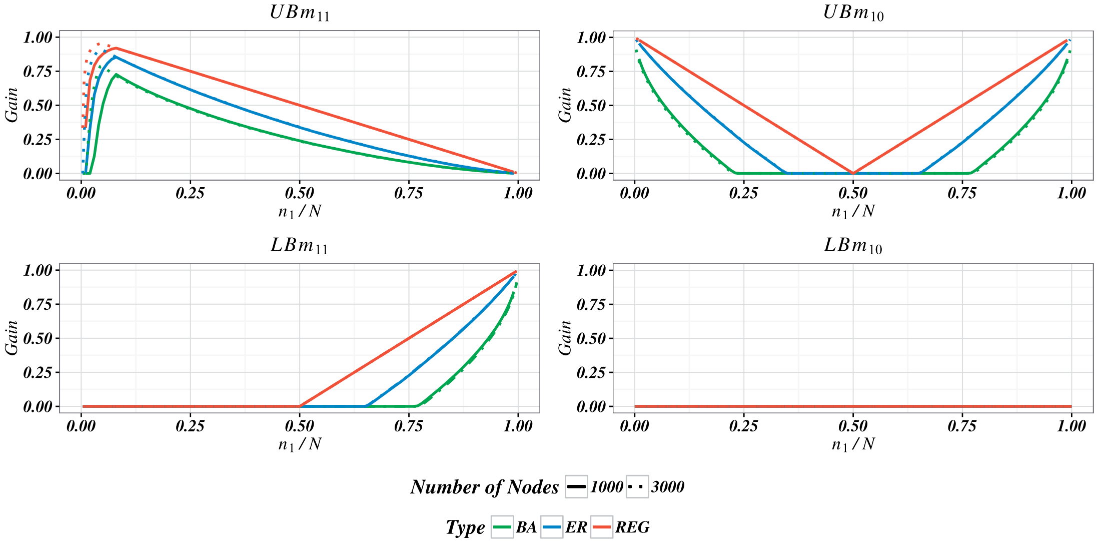

FIG. 10. Clockwise representation of the gain deriving from the application of $UBm_{11}$, $UBm_{10}$, $LBm_{11}$, $LBm_{10}$ as a function of $n_1$ nodes on $N = 1000, 3000$. Average degree was set to $\langle d \rangle = 6$.

### 4.2 *Analysis of benchmark instances*

We applied the proposed bounds to three classes of networks of various size according to their degree distribution choosing Erdős and Rényi random graphs ([12]), scale-free networks [1] and regular graphs. For the first two classes, we generated ten instances with same degree distribution, since the bounds provide identical results for each realization of the same degree sequence. Thus, with 10 different degree sequences we study the gain on a set of similar, but slightly different, $D_G$. When computing the gain, differences among instances result attenuated and, finally, the average try to catch the real behaviour of the system.

We considered networks with $N = 1000$ and 3000 with two different settings: the first with an average degree $\langle d \rangle = 6$; the second having a density $\delta = 0.9$. In addition, for each setting we included regular graphs. The first setting was considered in order to perform an analysis similar to [30] while the second setting was chosen in order to test bounds that require $D_G$ from dense graphs.

The curves in Fig.10 shows the percentage gain obtained from applying each new bound from the perspective of the area covered by the feasible region, with respect to the bounds in formula (2.5) and (2.6) when $\langle d \rangle = 6$. This kind of analysis allows us to estimate the bound behaviour and the dependency of the gain for different networks types. Indeed, we can firstly observe how, fixed the mean degree, the networks size do not affect the bounds as behave exactly the same, while the value of $n_1$ acts as a threshold to trigger the bounds. Indeed, upper bounds of formulas (3.1) and (3.2) have a non-homogeneous behavior and for some values of $n_1$ tend to get closer to the values achieved by the bound in formulas (2.5) and (2.6).

Through observing the analytical relationships referred the upper bounds, it is evident that homogeneous $D_G$ tends to perform better, while those of a more heterogeneous nature such as, for instance, scale-free networks, perform slightly worse due to the deviation of some nodes from the mean degree.

Lower bound $LBm_{11}$ evaluation can be performed using a similar approach as for the upper bounds. Note that we cannot make any comparisons with previous results being implicitly set to 0 in [30].

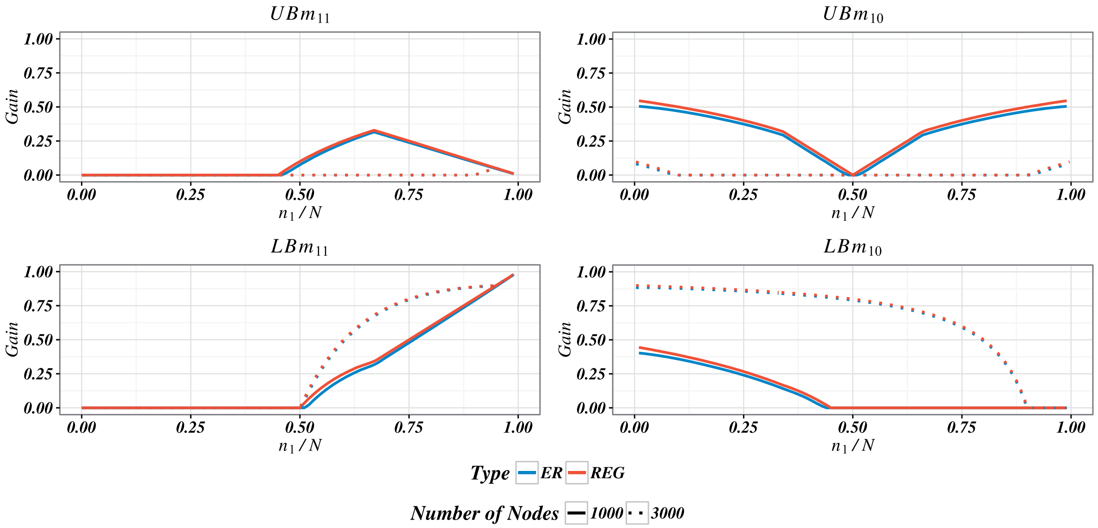

FIG. 11. Clockwise representation of the gain deriving from the application of $UBm_{11}$, $UBm_{10}$, $LBm_{11}$, $LBm_{10}$ as a function of $n_1$ nodes on $N = 1000, 3000$. Density was set to $\delta = 0.9$.

$LBm_{10}$ gives a contribution that can be considered close to zero due to the fact that it is computed as the difference between the tail of $D_G$ and the edges within $K_{n_1}$. Indeed, when $\langle d \rangle = 6$ we deal with very sparse graph, thus the latter difference is always negligible.

We tuned the density to high values ($\delta = 0.9$) since the terms in $LBm_{10}$ formula depend on high degree elements within $D_G$. Fig. 11 shows the gain obtained through the introduction of new bounds for both Erdős-Rényi and regular graphs. Scale-free networks were omitted in the test as they are characterized by low density values [9].

In this case, the impact of both lower bounds is more significant, particularly for $LBm_{10}$ which differs to zero most substantially when $N$ grows. Indeed, as we noticed before, all the bounds depend on the mean degree $\langle d \rangle$. Therefore, to keep the value of the density $\delta$ while increasing the network size, implies the growth of the mean degree as well, since $\langle d \rangle = \delta(N - 1)$. Comparing the curves in Figure 11 it is evident that when the value of $N$ is higher, the impact of the upper bounds tends to be more negligible since their gain is substituted by that of the lower bounds. Finally, the performed analysis reinforces the importance of introducing lower bounds, especially when dealing with dense graphs.

## 5. Conclusions

In this article, we developed two upper bounds and two lower bounds in order to reduce the area of the two dimensional space used to represent the feasible region of the dyadic effect in a network with certain nodes characteristics. Using commonly accepted structural principles, we improved the upper bounds and provided two new lower bounds. The four bounds can be computed using straightforward analytical relationships with no restrictions on either the network size or classes with the only limitation for the graph to be simple. The computational analysis of various classes of networks resulted in behavioural differences depending on the inner structure and on the shape of the degree sequence.

These results are particularly relevant in applications where large networks are investigated and numerous characteristics are taken into account as a way of making valid hypotheses on which has been more significant in determining the observed topology. Under these circumstances, the maximum and minimum values of dyadicity and heterophilicity act as a threshold for the computation of efficient and tight relative measures. This kind of approach is able to provide an informative content which represents a reasonable alternative to the one of a complete enumeration in a less onerous way.

Further research should be devoted to study additional aspects. In particular, the asymptotic behaviour of each bound when the network size grows should be considered. Another point of interest would be to improve the bounds using further arguments, especially in the case of scale-free networks where the proposed approach have seemed to perform less effectively.

Finally, considering the strict relationship between assortative mixing and dyadic effect, the proposed bounds and their implications could lay the bases for possible reformulations and improvements of the metrics related to the study of assortative mixing as well as for other potential studies in such direction. For instance, one could be interested in further investigating the nexus between the assortativity and the presence of different values of dyadicity and heterophilicity. Finally it would be interesting to study how the presence of large groups and communities of featured nodes shapes the phase diagram contained in the region determined by the new bounds that we introduced. Furthermore, the use of local substructures as triangles or motifs, together with a more detailed analysis of the neighbourhood of each node, could be investigated to improve the quality of our bounds. Taking into account the wide range of applications of the presented bounds it is worth to consider the trade-off between their usefulness and tightness. Indeed, future refinements should lead to tighter bounds exploiting topological aspects beyond the degree sequence still maintaining their applicability.

## Acknowledgments

The authors would like to thank Benedetto Scoppola for many useful conversations and comments.

## References

1. Barabási, A.-L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
2. Barrat, A., Barthelemy, M. & Vespignani, A. (2008) *Dynamical Processes on Complex Networks*. New York, USA: Cambridge University Press.
3. Bianconi, G., Pin, P. & Marsili, M. (2009) Assessing the relevance of node features for network structure. *Proc. Natl. Acad. Sci. USA*, **106**, 11433–11438.
4. Cinelli, M., Ferraro, G. & Iovanella, A. (2016) Some insights into the relevance of nodes' characteristics in complex network structures. *Designing Networks for Innovation and Improvisation, Proceedings of the 6th International COINs Conference* (M. P. Zylka, H. Fuehres, A. F. Colladon & P. A. Gloor eds). Cham, Switzerland: Springer Proceedings in Complexity, Springer International Publishing.
5. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. (2006) Detecting rich-club ordering in complex networks. *Nat. Phys.*, **2**, 110–115.
6. Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.
7. D'Agostino, G., Scala, A., Zlatić, V. & Caldarelli, G. (2012) Robustness and assortativity for diffusion-like processes in scale-free networks. *EPL (Europhys. Lett.)*, **97**, 68006.
8. de Almeida, L. M., Mendes, A. G., Madras Viswanathan, G. & da Silva, R. L. (2013) Scale-free homophilic network. *Eur. Phys. J. B*, **86**, 1–6.
9. Del Genio, C. I., Gross, T. & Bassler, K. E. (2011) All scale-free networks are sparse. *Phys. Rev. Lett.*, **107**, 178701.

**10.** Di Paola, L., De Ruvo, M., Paci, P., Santoni, D. & Giuliani, A. (2012) Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.*, **113**, 1598–1613.

**11.** Erdős, P. & Gallai, T. (1960) Graphs with prescribed degrees of vertices (in Hungarian). *Matematikai. Lapok*, **11**, 264–274.

**12.** Erdős, P. & Rényi, A. (1959) On random graphs. *Publicationes Mathematicae*, **6**, 290–297.

**13.** Estrada, E. (2011) Combinatorial study of degree assortativity in networks. *Phys. Rev. E*, **84**, 047101.

**14.** Ferraro, G. & Iovanella, A. (2015) Organizing collaboration in inter-organizational innovation networks, from orchestration to choreography. *Int. J. Eng. Business Manag.*, **7**, 24.

**15.** Ferraro, G. & Iovanella, A. (2016) Revealing correlations between structure and innovation attitude in inter-organisational innovation networks. *Int. J. Comput. Econ. Economet.*, **6**, 93–113.

**16.** Ferraro, G., Iovanella, A. & Pratesi, G. (2016) On the influence of nodes' characteristic in inter-organisational innovation networks structure. *Int. J. Comput. Econ. Economet.*, **6**, 239–257.

**17.** Garey, M. R. & Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: WH Freeman and Company.

**18.** Hakimi, S. L. (1962) On realizability of a set of integers as degrees of the vertices of a linear graph. I. *J. Soc. Ind. Appl. Math.*, **10**, 496–506.

**19.** Havel, V. (1955) A remark on the existence of finite graphs. *Casopis Pest. Mat.*, **80**, 1253.

**20.** Hu, T., Andrew, A. S., Karagas, M. R. & Moore, J. H. (2015) Functional dyadicity and heterophilicity of gene-gene interactions in statistical epistasis networks. *BioData Mining*, **8**, 1.

**21.** Jiang, J. Q., Dress, A. & Chen, M. (2010) Towards prediction and prioritization of disease genes by the modularity of human phenome-genome assembled network. *J. Integr. Bioinform.*, **7**, 149.

**22.** Jiang, X., Liu, B., Jiang, J., Zhao, H., Fan, M., Zhang, J., Fan, Z. & Jiang, T. (2008) Modularity in the genetic disease-phenotype network. *FEBS Lett.*, **582**, 2549–2554.

**23.** Kim, H., Toroczkai, Z., Erdős, P. L., Miklós, I. & Székely, L. A. (2009) Degree-based graph construction. *J. Phys. A Math. Theor.*, **42**, 392001.

**24.** Liu, D., Trajanovski, S. & Van Mieghem, P. (2013) Random line graphs and a linear law for assortativity. *Phys. Rev. E*, **87**, 012816.

**25.** McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001) Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.*, **27**, 415–444.

**26.** Newman, M., Barabási, A.-L. & Watts, D. J. (2006) *The Structure and Dynamics of Networks: (Princeton Studies in Complexity)*. Princeton, NJ: Princeton University Press.

**27.** Newman, M. E. (2003) Mixing patterns in networks. *Phys. Rev. E*, **67**, 026126.

**28.** Newman, M. E., Watts, D. J. & Strogatz, S. H. Random graph models of social networks. *Proc. Natl. Acad. Sci. USA*, **99**, 2566–2572.

**29.** Noldus, R. & Van Mieghem, P. (2015) Assortativity in complex networks. *J. Complex Netw.*, **3**, 507.

**30.** Park, J. & Barabási, A.-L. (2007) Distribution of node characteristics in complex networks. *Proc. Natl. Acad. Sci. USA*, **104**, 17916–17920.

**31.** R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

**32.** van den Heuvel, M. P., Kahn, R. S., Goñi, J. & Sporns, O. (2012) High-cost, high-capacity backbone for global brain communication. *Proc. Natl. Acad. Sci. USA*, **109**, 11372–11377.

**33.** White, D. R. & Harary, F. (2001) The cohesiveness of blocks in social networks: node connectivity and conditional density. *Sociol. Methodol.*, **34**, 305–359.

**34.** Zhang, X., Zhang, R., Jiang, Y., Sun, P., Tang, G., Wang, X., Lv, H. & Li, X. (2011) The expanded human disease network combining protein–protein interaction information. *Eur. J. Hum. Genet.*, **19**, 783–788.

**35.** Zhou, D., Stanley, H. E., DâŁ™Agostino, G. & Scala, A. (2012) Assortativity decreases the robustness of interdependent networks. *Phys. Rev. E*, **86**, 066103.

**36.** Zhou, S. & Mondragón, R. J. (2004) The rich-club phenomenon in the Internet topology. *IEEE Commun. Lett.*, **8**, 180–182.