BIOINFORMATICS

Vol. 00 no. 00 2005
Pages 1–9

Using coalitional games on biological networks to measure centrality and power of genes

Stefano Moretti [1,2,*], Vito Fragnelli [3], Fioravante Patrone [4] and Stefano Bonassi [5]

[1] CNRS, UMR7024, F-75016 Paris, France.
[2] Université Paris-Dauphine, Lamsade, F-75016 Paris, France. email: Stefano.MORETTI@dauphine.fr
[3] Department of Advanced Sciences and Technologies, University of Eastern Piedmont, Italy. email: vito.fragnelli@mfn.unipmn.it
[4] DIPTEM, University of Genova, Italy. email: patrone@diptem.unige.it
[5] Unit of Clinical and Molecular Epidemiology, IRCCS San Raffaele Pisana, Roma, Italy. email: stefano.bonassi@sanraffaele.it

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: The interpretation of gene interaction in biological networks, generates the need for a meaningful ranking of network elements. Classical centrality analysis ranks network elements according to their importance but may fail to reflect the power of each gene in interaction with the others.

Results: We introduce a new approach using coalitional games to evaluate the centrality of genes in networks keeping into account genes interactions. The Shapley value for coalitional games is used to express the power of each gene in interaction with the others and to stress the centrality of certain hub genes in the regulation of biological pathways of interest. The main advantage of games on interaction networks, with respect to previous applications of game theory to gene expression analysis, is a finer resolution of the gene interaction investigated in the model, which is based on pair-wise relationships of genes in the network. In addition, the new approach allows for the integration of *a priori* knowledge about genes playing a key function on a certain biological process.

An approximation method for practical computation on large biological networks, together with a comparison with other centrality measures, is also presented.

Contact: Stefano.MORETTI@dauphine.fr

1 INTRODUCTION

Gene expression data may be collected by means of microarray technology (Golub *et al*. (1999); Parmigiani *et al*. (2003)). Within a single experiment of this sophisticated technology, the level of expression of thousands of genes is estimated in a sample of cells under given conditions (genetic diseases, environmental exposition, pharmacologic treatment, levels of activation of a given pathway of genes etc.). Several approaches have been proposed to identify "central" genes of different biological pathways within the huge amount of information provided by this technology (Amaratunga and Cabrera (2004); Tusher *et al*. (2001); Storey and Tibshirani (2003)).

Gene co-expression networks (Zhang and Horvath (2005)) and other biological networks (e.g. representing protein-protein interactions) are increasingly used to explore the system-level functionality of genes and proteins (Carlson *et al*. (2006); Jeong *et al*. (2001)). Co-expression networks, for instance, are connection situations based upon the extent of correlation between pairs of genes across a gene expression dataset. Nodes are genes and connections are defined by co-expression of two genes. Often, the Pearson correlation coefficient is the initial measure of gene co-expression. This measure is then transformed into an adjacency matrix, according to different alternative statistical procedures (Zhang and Horvath (2005); Carlson *et al*. (2006)). Depending on the aims of the study, weighted or un-weighted networks, generated by the dichotomization of the corresponding correlation matrix, may be considered. The need of interpreting gene interaction in co-expression networks requires the ranking of network elements. Centrality analysis ranks single elements according to their importance within the network structure, and different measures of centrality focus on various aspects of the structure of a network (Mason and Verwoerd (2007); Junker *et al*. (2006)), e.g., most central elements of protein networks were essential to predict lethal mutations (Jeong *et al*. (2001)). Highly connected hub genes, largely responsible for maintaining network connectivity, were likely essential for yeast survival (Carlson *et al*. (2006)), although standard centrality measures may fail to reflect the power of each gene to interact with the others.

Cooperative game theory may also be used to analyze gene expression data (see for instance Albino *et al*. (2008); Fragnelli and Moretti (2008); Jeong *et al*. (2001); Lucchetti *et al*. (2009); Moretti *et al*. (2007, 2008); Moretti (2009, 2010)). In Moretti *et al*. (2007) the class of microarray games has been introduced to quantitatively evaluate the relevance of each gene in generating

*to whom correspondence should be addressed

or regulating a condition of interest (e.g. a disease), taking into account the observed relationships in all subgroups of genes. In the framework of microarray games, the relevance of genes is expressed in terms of the Shapley value (Shapley (1953); Moretti and Patrone (2008)). The Shapley value attributed to a certain gene in a given microarray game corresponds to the relevance of that gene for the mechanisms governing the genomic effects of the condition under study. This game theoretic approach has been successfully applied to real datasets (Albino *et al*. (2008); Moretti *et al*. (2008)) and provides a characterization of a relevance index for genes which is mainly based on the role they play inside gene-regulatory pathways (Moretti *et al*. (2007)). A comparison between the results provided by the analysis of the Shapley value of microarray games and the results provided by classic statistical testing is discussed in connection with the pathogenesis of neuroblastic tumors in Albino *et al*. (2008), and in Moretti *et al*. (2008), where gene expression in children differentially exposed to air pollution is studied.

Standard centrality measures (Mason and Verwoerd (2007); Junker *et al*. (2006)) do not take into account the strength of interrelations inside subgroups of genes, in contrast with a central issue of coalitional games in cooperative game theory, which is precisely to analyze the overall power of players according to their role in all feasible "coalitions" (Shapley (1953); Moretti and Patrone (2008)). In the context of social networks, Gòmez *et al*. (2003) proposed a new family of centrality measures based on coalitional games defined on networks, i.e. measures of the importance of nodes in a network where links reflect the interactions among individuals (nodes). Our idea was to use a similar approach in the context of co-expression networks. We define an association game as a coalitional game (also known as a cooperative game in characteristic function form) (N, v), where N is the set of genes studied in the expression dataset and v is the characteristic function, which assigns a "worth" to each subset (coalition) of genes in N. The worth of a coalition represents the overall magnitude of the correlation between the genes of the coalition and a set of key-genes selected *a priori* (e.g. a set of genes known to be involved in biological pathways related to chromosome damage).

In order to study the cascade of activation/deactivation among genes, gene interaction is restricted to the connections within an associated interaction network or co-expression network Γ, and therefore another coalitional game (N, w^v_Γ) is studied, which is defined as the restriction Myerson (1977) of the association game (N, v) to the co-expression network computed on the dataset. The difference of the Shapley values computed on the two coalitional games (N, v) and (N, w^v_Γ) is considered as a gene centrality measure.

The paper is organized as follows. Next section, after the introduction of some preliminary notations, is devoted to the presentation of the game theoretic centrality measure. Section 3 presents a preliminary application of the method to a real dataset. Section 4 introduces an approximation method for centrality computation and the comparison of the results with other centrality measures on a large network. Section 5 concludes.

2 APPROACH

2.1 Preliminaries

An (undirected) *graph* or *network* is a pair $\langle V, E\rangle$, where V is a set of vertices or nodes and E is a set of edges e of the form $\{i, j\}$ with $i, j \in V$, $i \neq j$.

A *path* between i and j in a graph $\langle V, E\rangle$ is a sequence of nodes $(i_0, i_1, \ldots, i_k)$, where $i = i_0$ and $j = i_k$, $k \geq 1$, such that $\{i_s, i_{s+1}\} \in E$ for each $s \in \{0, ..., k-1\}$ and such that all these edges are distinct. Two nodes $i, j \in V$ are connected in $\langle V, E\rangle$ if $i = j$ or if there exists a path between i and j in E.

A *cycle* in $\langle V, E\rangle$ is a path from i to i for some $i \in V$. A path $(i_0, i_1, \ldots, i_k)$ is *without cycles* if there do not exist $a, b \in \{0, 1, \ldots, k\}$, $a \neq b$, such that $i_a = i_b$. A *forest* is a graph where each path is without cycles.

A *connected component* of V in $\langle V, E\rangle$ is a maximal subset of V with the property that any two nodes in this subset are connected in $\langle V, E\rangle$. The set of all the connected components in $\langle V, E\rangle$ is denoted by $\mathcal{C}_E$.

Now, we introduce some basic game theoretical notations. A *coalitional game* or *characteristic-form game* is a pair (N, v), where N denotes the finite set of *players* and $v : 2^N \to I\!R$ the *characteristic function*, with $v(\emptyset) = 0$. If the set N of players is fixed, we identify a coalitional game (N, v) with the corresponding characteristic function v. A group of players $T \subseteq N$ is called a *coalition* and $v(T)$ is called the *worth* of this coalition. We will denote by $\mathcal{G}$ the class of all coalitional games.

Let $\mathcal{C} \subseteq \mathcal{G}$ be a subclass of coalitional games. Given a set of players N, we denote by $\mathcal{C}^N \subseteq \mathcal{C}$ the class of coalitional games in $\mathcal{C}$ with N as set of players.

The *unanimity game* (N, u_R) on $R \subseteq N$ is the game described by $u_R(T) = 1$ if $R \subseteq T$ and $u_R(T) = 0$, otherwise. Every coalitional game (N, v) can be written as a linear combination of unanimity games in a unique way, *i.e.* $v = \sum_{S \subseteq N, S \neq \emptyset} \lambda_S(v) u_S$ (see, for instance, Owen (1995)). The coefficients $\lambda_S(v)$, for each $S \in 2^N \setminus \{\emptyset\}$, are called *unanimity coefficients* of the game (N, v).

Let $|N|$ be the cardinality of a finite set N. A *payoff vector* or *allocation* $(x_1, \ldots, x_n)$ of a coalitional game (N, v) is an $|N|$-dimensional vector describing the payoffs of the players, such that each player $i \in N$ receives x_i.

A *one-point solution* (or simply a *solution*) for a class $\mathcal{C}^N$ of coalitional games is a function ψ that assigns a payoff vector $\psi(v)$ to every coalitional game in the class, that is $\psi : \mathcal{C}^N \to I\!R^N$.

The most widely used solution in the theory of coalitional games is the *Shapley value*, introduced by Shapley in 1953 (Shapley (1953)). This solution can be described in several ways. In order to provide its original definition, we first need to introduce the notions of order on N and of marginal vector.

We define the set Σ_N of possible orders on the set N as the set of all bijections $\sigma : \{1, \ldots, |N|\} \to N$, where $|N|$ is the cardinality of the set N and where $\sigma(i) = j$ means that with respect to σ, player j is in the i-th position. Let (N, v) be a coalitional game with N as the set of players. For $\sigma \in \Sigma_N$, the *marginal vector* $m^\sigma(v)$ is defined by

$$m^\sigma_i(v) = v([i, \sigma]) - v((i, \sigma)) \text{ for all } i \in N,$$

where $[i, \sigma] = \{j \in N : \sigma^{-1}(j) \leq \sigma^{-1}(i)\}$ is the set of predecessors of i with respect to σ including i, and $(i, \sigma) = \{j \in$

$N : \sigma^{-1}(j) < \sigma^{-1}(i)\}$ is the set of predecessors of i with respect to σ excluding i.

The Shapley value $\phi(v)$ of a game (N, v) is then defined as the average of marginal vectors over all $|N|!$ possible orders in Σ_N. In formula

$$\phi_i(v) = \sum_{\sigma \in \Sigma_N} \frac{m_i^\sigma(v)}{|N|!} \text{ for all } i \in N. \quad (1)$$

An alternative representation of the Shapley value can be given in terms of the unanimity coefficients $(\lambda_S(v))_{S \in 2^N \setminus \{\emptyset\}}$ of a game (N, v), that is:

$$\phi_i(v) = \sum_{S \subseteq N : i \in S} \frac{\lambda_S(v)}{|S|} \quad (2)$$

for each $i \in N$.

2.2 Genes and games

Suppose to have a set K of *key-genes* assumed to be *equally important* for the regulation of a certain biological process. Let N be the set of genes who are studied on a sequence of (microarray) experiments under a condition of interest, for instance a genetic disorder. Let $I \subseteq \{\{i, k\} | i \in N, k \in K\}$ be the set of *interactions* between genes in N and key-genes in K. We will say that a gene $i \in N$ and a key-gene $k \in K$ *interact* if and only if $\{i, k\} \in I$. The triple (N, K, I) is said a *gene-k-gene* (gkg) situation.

Given a set of genes $S \subseteq N$, the higher the number of key-genes which interact with genes in S, the higher the likelihood that genes in S are also involved in the regulation of the biological process of interest. In order to measure the strength of *association* of pathways of genes in N, for each group $S \subseteq N$ we compute the number of key-genes interacting only with genes in S. Let $v : 2^N \to \mathbb{N}$ be the map assigning to each coalition $S \in 2^N \setminus \{\emptyset\}$ the number $v(S)$ of key-genes in K which only interact (in I) with genes in S. By convention, $v(\emptyset) = 0$. The pair (N, v) is called *association game* corresponding to (N, K, I). Note that the assumption of equal importance for key-genes is central for the definition of the characteristic function v. In fact, the value $v(S)$, for each $S \in 2^N \setminus \{\emptyset\}$, represents a measure of the relevance of coalition S in terms of the number of key-genes directly interacting only with genes in S. The possibility to compare the relevance of different coalitions makes sense thanks to the assumption of equal importance of key-genes.

In the remaining of the paper, to simplify the presentation of the game theoretic model, we will also assume that key-genes are *independent*, i.e. they do not directly interact between them. However, this assumption is not fundamental as the one of equal importance. If a group of m key-genes directly interact, it will be sufficient to collapse them into an individual key-unit whose importance equals m times the importance of a single key-genes.

EXAMPLE 1. *Consider a set of genes $N = \{1, 2, 3, 4\}$, a set of key-genes $K = \{a, b, c\}$ and a set of interactions $I = \{\{1, a\}, \{1, b\}, \{3, b\}, \{3, c\}, \{4, c\}\}$. The association game (N, v) is such that $v(\emptyset) = v(2) = v(3) = v(4) = v(2, 3) = v(2, 4) = 0$, $v(1, 3) = v(1, 2, 3) = 2$, $v(1, 3, 4) = v(1, 2, 3, 4) = 3$ and $v(S) = 1$ for all the remaining coalitions.*

If gene $i \in N$ has not directly an interaction with $k \in K$, it may still be possible for i to interact with k via an interaction with another gene $j \in N$ (an intermediary) which in turn has an interaction with k, or more generally, via a sequence of intermediaries. So, it is essential to understand which genes really interact, directly or via intermediaries, and how the network of such interactions may affect the worth of coalitions of genes.

Let us consider now an *interaction network* $\langle N, \Gamma \rangle$, the nodes of the graph being the genes. The set of edges Γ indicates interaction ties between pairs of genes, i.e. a set $\{i, j\} \subseteq N$ is an element of Γ if and only if i and j have an interaction. Implicitly, this graph shows us which coalitions are feasible, i.e., which coalitions have all their members related by interactions.

Given a gkg situation (N, K, I) with the corresponding association game (N, v) and an interaction network $\langle N, \Gamma \rangle$, following the approach in Myerson (1977), we use the structure of an interaction network to define a new game (N, w_Γ^v), where the value $w_\Gamma^v(S)$ of a coalition S equals the sum of the values assigned by v to the connected components of the network restricted to this coalition S. The game w is called the graph-restricted game.

DEFINITION 1. Let (N, K, I) be a gkg situation and let (N, v) be the corresponding association game. Let $\langle N, \Gamma \rangle$ be an interaction network. The *graph-restricted game* (N, w_Γ^v) is defined by

$$w_\Gamma^v(S) = \sum_{T \in C_{\Gamma_S}} v(T) \quad (3)$$

for each $S \in 2^N \setminus \{\emptyset\}$, where C_{Γ_S} is the set of all the connected components in $\langle S, \Gamma_S \rangle$, and with the convention $w_\Gamma^v(\emptyset) = 0$.

EXAMPLE 2. *Consider the gkg situation of Example 1 with the corresponding association game (N, v). Consider the interaction network $(N, \hat{\Gamma})$ where $\hat{\Gamma} = \{\{1, 2\}, \{2, 3\}\}$. All the interactions are represented in network of Figure 1.*

The graph-restricted game $(N, w_{\hat{\Gamma}}^v)$ is such that $w_{\hat{\Gamma}}^v(1) = w_{\hat{\Gamma}}^v(1, 2) = w_{\hat{\Gamma}}^v(1, 4) = w_{\hat{\Gamma}}^v(1, 2, 4) = w_{\hat{\Gamma}}^v(1, 3) = w_{\hat{\Gamma}}^v(1, 3, 4) = 1$, $w_{\hat{\Gamma}}^v(1, 2, 3) = 2$, $w_{\hat{\Gamma}}^v(1, 2, 3, 4) = 2$ and $w_{\hat{\Gamma}}^v(S) = 0$ for all the remaining coalitions.

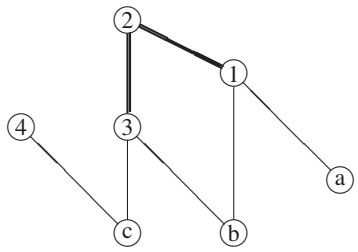


Fig. 1. Interaction network $\hat{\Gamma}$ (thick lines) and the interactions of the gkg situation described in Example 1 (thin lines).

EXAMPLE 3. *Consider the gkg situation of Example 1 with the corresponding association game (N, v). Consider the interaction network $(N, \bar{\Gamma})$ where $\bar{\Gamma} = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$. All the interactions are represented in network of Figure 2.*

The graph-restricted game $(N, w_{\bar{\Gamma}}^v)$ is such that $w_{\bar{\Gamma}}^v(3, 4) = w_{\bar{\Gamma}}^v(2, 3, 4) = 1$, $w_{\bar{\Gamma}}^v(1) = w_{\bar{\Gamma}}^v(1, 2) w_{\bar{\Gamma}}^v(1, 4) = w_{\bar{\Gamma}}^v(1, 2, 4) = w_{\bar{\Gamma}}^v(1, 3) = 1$, $w_{\bar{\Gamma}}^v(1, 2, 3) = w_{\bar{\Gamma}}^v(1, 3, 4) = 2$, $w_{\bar{\Gamma}}^v(1, 2, 3, 4) = 3$ and $w_{\bar{\Gamma}}^v(S) = 0$ for all the remaining coalitions.

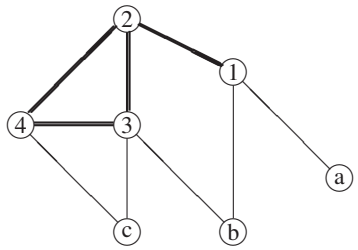


Fig. 2. Interaction network $\bar{\Gamma}$ (thick lines) and the interactions of the gkg situation described in Example 1 (thin lines).

Since the basic paper Shapley and Shubik (1954), in several different applications the Shapley value of a game has been considered as a player's power (see, for instance, the survey Moretti and Patrone (2008) for references to the use of the Shapley value as a power index in different contexts). Here, players are genes and the Shapley value is considered as a gene's power. The intuition behind the meaning of gene's power attributed to relation (1) follows from this consideration. An order σ on a set of genes N may be interpreted as a sequence of activations of study-genes and the corresponding marginal vector may be seen as a measure of the power of study-genes to establish relevant interactions with key-genes according to σ. However, in absence of information about which sequences of activations are more likely, it is reasonable to average the marginal vectors over all possible orders as an indication of the expected power of genes.

The difference between the power of a gene in the graph-restricted game and its power in the association one is proposed as a centrality measure for co-expression networks (see Gòmez *et al*. (2003) in the context of social networks). Let (N, K, I) be a gkg situation and let (N, v) be the corresponding association game. Let $\langle N, \Gamma\rangle$ be an interaction network. The *centrality measure* $\gamma(v, \Gamma)$ is defined by

$$\gamma_i(v, \Gamma) = \phi_i(w^v_\Gamma) - \phi_i(v), \tag{4}$$

for each $i \in N$, where $\phi(v)$ is the Shapley value of the association game v and $\phi(w^v)$ is the Shapley value of the corresponding graph-restricted game w^v_Γ.

EXAMPLE 4. *Consider the gkg situation with the corresponding association game (N, v) and the interaction network of Example 2. According to relation (1), we have that $\phi(v) = (\frac{3}{2}, 0, 1, \frac{1}{2})$ and $\phi(w^v_{\hat{\Gamma}}) = (\frac{4}{3}, \frac{1}{3}, \frac{1}{3}, 0)$. Thus, the centrality measure gives $\gamma(v, \hat{\Gamma}) = (-\frac{1}{6}, \frac{1}{3}, -\frac{2}{3}, -\frac{1}{2})$.*

As an example of Shapley value computation via relation (1), we show here the calculation for gene 1. In total there are $4! = 24$ orders in Σ_N. There are precisely 6 orders $\sigma \in \Sigma_N$ such that $\sigma^{-1}(1) = 1$ and other 6 orders $\sigma \in \Sigma_N$ such that $\sigma^{-1}(4) = 1$. In addition, for each intermediate coalition $S \subseteq \{2, 3, 4\}$ of one or two genes, there are two orders on N such that S is the set of precessors of 1. Consequently, from relation (1), the Shapley value of gene 1 is

$$\begin{aligned}
&\phi_1(v) = \tfrac{1}{24}\big(6(v(\{1) - v(\emptyset)) + 6(v(1,2,3,4) - v(2,3,4))\\
&+2(v(1,2) - v(2)) + 2(v(1,3) - v(3)) + 2(v(1,4) - v(4))\\
&+2(v(1,2,3) - v(2,3)) + 2(v(1,3,4) - v(3,4))\\
&+2(v(1,2,4) - v(2,4))\big)\\
&= \tfrac{1}{24}\big(6 \times (1-0) + 6 \times (3-1) + 2 \times (1-0) + 2 \times (2-0)\\
&+2 \times (1-0) + 2 \times (2-0) + 2 \times (3-1) + 2 \times (1-0)\big)\\
&= \tfrac{1}{24}\big(6 + 12 + 2 + 4 + 2 + 4 + 4 + 2\big) = \tfrac{36}{24} = \tfrac{3}{2}.
\end{aligned}$$

Next section is devoted to illustrate a more efficient way to calculate the Shapley value of genes.

EXAMPLE 5. *Consider the gkg situation with the corresponding association game (N, v) and the interaction network of Example 3. Again, according to relation (1), we have that $\phi(v) = (\frac{3}{2}, 0, 1, \frac{1}{2})$ (nothing changed in game v) and $\phi(w^v_\Gamma) = (\frac{4}{3}, \frac{1}{3}, \frac{5}{6}, \frac{1}{2})$. Thus, the centrality measure gives $\gamma(v, \bar{\Gamma}) = (-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, 0)$. Note that with respect to Example 4, where edges $\{2, 4\}$ and $\{3, 4\}$ were not present, gene 2 continues to be the unique one with strictly positive centrality according to γ, even if genes 3 and 4 increase their centrality.*

It should be noted that $-v(N) < \gamma_i(v, \Gamma) < w^v_\Gamma(N)$ for each $i \in N$. As a consequence, γ centrality computed on different interaction networks are comparable scores only if they are defined on the same interval scale, that is if the worth of the largest coalition in the graph-restricted game is the same for both interaction networks.

2.3 Centrality computation

Actually, the computation of the Shapley value using relation (1) may be very hard even if the number of genes is quite small. For instance, note that with only 10 genes it necessary to consider $10! = 3628800$ orders of genes in (1). In order to make real applications, it is useful to decompose the association game and the corresponding graph-restricted game by means of a relatively small number of unanimity games with non-null unanimity coefficients. As a consequence, the Shapley value of such games may be computed in a less complex way via relation (2). In the following we briefly describe this decomposition procedure.

Let (N, K, I) a gkg situation. For each key-gene $k \in K$, the set of genes in N which have a strong interaction with k are denoted by $N_k = \{i \in N | \{i, k\} \in I\}$. Let (N, v) the corresponding association game. It is easy to show that the characteristic function v can be written as a sum of unanimity games:

$$v = \sum_{k \in K, N_k \neq \emptyset} u_{N_k}. \tag{5}$$

EXAMPLE 6. *Consider the gkg of Example 1. We have that $N_a = \{1\}, N_b = \{1, 3\}, N_c = \{3, 4\}$. From relation (5), the corresponding association game v is given by*

$$v = u_{\{1\}} + u_{\{1,3\}} + u_{\{3,4\}}.$$

Consequently, according to relation (2), the Shapley value of v can easily be calculated as the following sum of vectors

$$\phi(v) = (1, 0, 0, 0) + (\tfrac{1}{2}, 0, \tfrac{1}{2}, 0) + (0, 0, \tfrac{1}{2}, \tfrac{1}{2}) = (\tfrac{3}{2}, 0, 1, \tfrac{1}{2}).$$

The remaining of this section is devoted to provide a natural decomposition of a graph-restricted game based on the reformulation of the association game given in (5).

First, we need to introduce the concept of minimal component containing a coalition S. Let $\langle N, E\rangle$ be a graph. We denote by $\langle N, F_E\rangle$ a graph where F_E is a maximal subset of E with the property that $\langle N, F_E\rangle$ is a forest. The set of all the forests for E is denoted by $\mathcal{F}_E$. Let $S \in 2^N \backslash \{\emptyset\}$. A *minimal component containing* S in a forest $\langle N, E\rangle$ is a minimal subset of N which contains S and with the property that any two nodes in this set are connected in $\langle N, E\rangle$. Note that in a forest, a minimal component containing S, if exists, is unique. This fact allows us to denote the minimal component containing S in a forest F_E (if it exists) by $M_{F_E}(S)$, and the set of all the minimal components containing S in a graph $\langle N, E\rangle$ is denoted by $\mathcal{M}_E(S) = \{M_{F_E}(S) | F_E \in \mathcal{F}_E\}$.

Let $\langle N, \Gamma\rangle$ be a graph. Consider an unanimity game (N, u_S), with $S \in 2^N \setminus \{\emptyset\}$ and such that $\mathcal{M}_\Gamma(S) \neq \emptyset$. Without loss of generality, suppose that $\mathcal{M}_\Gamma(S) = \{M_\Gamma^{i_1}(S), \ldots, M_\Gamma^{i_r}(S)\}$, with $r \geq 1$. We define a new game $(N, w_\Gamma^{u_S})$ in the following way

$$w_\Gamma^{u_S} = \sum_{j=1}^{r} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq r} u_{M_\Gamma^{i_1}(S) \cup \ldots \cup M_\Gamma^{i_j}(S)}. \quad (6)$$

EXAMPLE 7. *Consider the interaction network of Example 3. Let* $S = \{1, 3\}$. *Note that* $\mathcal{M}_\Gamma(S) = \{\{1, 2, 3\}, \{1, 2, 3, 4\}\}$. *From relation (6) we have that*

$$w_\Gamma^{u_S} = u_{\{1,2,3\}} + u_{\{1,2,3,4\}} - u_{\{1,2,3\} \cup \{1,2,3,4\}} = u_{\{1,2,3\}}.$$

Games defined according to relation (6) are crucial for the computation of the Shapley value of graph-restricted games in practical situations. In fact, it can be proved that the game $w_\Gamma^{u_S}$ is the restriction of the unanimity game u_S to graph Γ,and is also known as the connecting S in Γ game (Gòmez *et al*. (2004)). Consequently, it can be easily shown that given a gkg situation (N, K, I) with the corresponding association game (N, v) and an interaction network $\langle N, \Gamma\rangle$, the graph-restricted game (N, w_Γ^v) may be computed via the following formula

$$\begin{array}{l} w_\Gamma^v = \\ = \sum_{k \in K, N_k \neq \emptyset, \mathcal{M}_\Gamma(N_k) \neq \emptyset} w_\Gamma^{u_{N_k}} \\ = \sum_{k \in K, N_k \neq \emptyset, \mathcal{M}_\Gamma(N_k) \neq \emptyset} \\ \sum_{j=1}^{|\mathcal{M}_\Gamma(N_k)|} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq r} u_{M_\Gamma^{i_1}(N_k) \cup \ldots \cup M_\Gamma^{i_j}(N_k)}, \end{array} \quad (7)$$

where $\mathcal{M}_\Gamma(N_k) = \{M_\Gamma^{i_1}(N_k), \ldots, M_\Gamma^{|\mathcal{M}_\Gamma(N_k)|}(N_k)\}$, for each $k \in K$ with $N_k \neq \emptyset$ and $\mathcal{M}_\Gamma(N_k) \neq \emptyset$.

From relations (2) and (7), it immediately follows that the Shapley value of a graph-restricted game w_Γ^v can be computed using the following relation

$$\begin{array}{l} \phi_i(w_\Gamma^v) \\ = \sum_{k \in K, i \in N_k, \mathcal{M}_\Gamma(N_k) \neq \emptyset} \\ \sum_{j=1}^{|\mathcal{M}_\Gamma(N_k)|} (-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq r} \frac{1}{M_\Gamma^{i_1}(N_k) \cup \ldots \cup M_\Gamma^{i_j}(N_k)}, \end{array} \quad (8)$$

for each $i \in N$.

EXAMPLE 8. *Consider the gkg with the corresponding association game* (N, v) *and the interaction network of Example 2. Note that* $\mathcal{M}_\Gamma(N_a) = \{\{1\}\}$, $\mathcal{M}_\Gamma(N_b) = \{\{1, 2, 3\}\}$ *and* $\mathcal{M}_\Gamma(N_c) = \{\emptyset\}$.

According to relation (7), we can write the graph-restricted game w_Γ^v *as a sum of unanimity games*

$$w_\Gamma^v = u_{\{1\}} + u_{\{1,2,3\}}. \quad (9)$$

Consequently, $\phi(w_\Gamma^v) = (\frac{4}{3}, \frac{1}{3}, \frac{1}{3}, 0)$.

EXAMPLE 9. *Consider the gkg with the corresponding association game* (N, v) *and the interaction network of Example 3. Note that* $\mathcal{M}_\Gamma(N_a) = \{\{1\}\}$, $\mathcal{M}_\Gamma(N_b) = \{\{1, 2, 3\}, \{1, 2, 3, 4\}\}$ *and* $\mathcal{M}_\Gamma(N_c) = \{\{3, 4\}, \{2, 3, 4\}\}$.

According to relation (7), we can write the graph-restricted game w_Γ^v *as a sum of unanimity games*

$$\begin{array}{l} w_\Gamma^v = u_{\{1\}} \\ + u_{\{1,2,3\}} + u_{\{1,2,3,4\}} - u_{\{1,2,3\} \cup \{1,2,3,4\}} \\ + u_{\{3,4\}} + u_{\{2,3,4\}} - u_{\{3,4\} \cup \{2,3,4\}} \\ = u_{\{1\}} + u_{\{1,2,3\}} + u_{\{3,4\}}. \end{array} \quad (10)$$

Consequently, $\phi(w_\Gamma^v) = (\frac{8}{6}, \frac{2}{6}, \frac{5}{6}, \frac{3}{6})$.

In the next section, we present an application of this centrality measure on a microarray data from children exposed to air pollution (Moretti *et al*. (2008)).

3 PRELIMINARY APPLICATION

We present a preliminary application of the method to gene expression data published in van Leeuwen *et al*. (2008), where genome-wide oligonucleotide microarray analysis was applied to blood cells of 23 children from Teplice (TP) region in the Czech Republic. The TP region is a mining district characterized by high levels of airborne pollutants including carcinogens. We consider the gene expression matrix $\mathbf{X}$ of 20130 genes and 23 samples from TP that was distilled from the data filtering and preparation as described in van Leeuwen *et al*. (2008).

As a set of key-genes, we used four genes known to be strongly associated with micronuclei frequencies, a bio-marker of chromosome damage: 1) PRC1 (protein regulator of cytokinesis 1); 2) TP53 (tumor protein p53 (li-fraumeni syndrome)); 3) ZWINT (zw10 interactor); 4) CCNB2 (cyclin b2) (Figure 3, green nodes). As a first filtering step, absolute values of Pearson correlation coefficients between each study-gene and each key-gene were computed, providing four lists of correlation coefficients (one list for each key-gene) with 20130 genes each, and the union of the top 25 genes from the four lists were selected for further analysis ($n = 96$). From the gene expressions of the selected 96 genes the corresponding gene correlation matrix was computed, and an unweighted network, based on dichotomizing the correlation matrix, was considered. More precisely, two genes were considered to interact (i.e. linked by an edge in the network) if and only if their absolute Pearson correlation coefficient was greater than 0.75 (Figure 3).

According to this criterion, it was possible to define the association game on the total set of 100 genes as the set of players, and the corresponding graph-restricted game. From the association game, only 9 genes obtained a non-null Shapley value ranging from

1 to 0.25 (Figure 3, yellow nodes). In fact, from relation (5), the association game is defined by

$$v = u_{\{1,2,3,4\}} + u_{\{26,27,28\}} + u_{\{38\}} + u_{\{72\}}$$

and, as a consequence of relation (2), $\phi_1(v) = \phi_2(v) = \phi_3(v) = \phi_4(v) = 0.25$, $\phi_{26}(v) = \phi_{27}(v) = \phi_{28}(v) = \frac{1}{3}$, $\phi_{38}(v) = \phi_{72}(v) = 1$ and $\phi_i(v) = 0$ for each other gene i.

In order to compute the Shapley value on the graph-restricted game, as it was described in Section 2.3, we first should find the sets of minimal connected components $\mathcal{M}_\Gamma(N_k)$, for each key-gene k. By definition, this requires the computation of the set of all the forests $\mathcal{F}(\Gamma)$. Several algorithms exist for generating all spanning trees of a graph that can be easily adapted to find all the forests (e.g., Gabow and Myers, (1978); Kapoor and Ramesh, (1995); Minty (1965)). However, as the number of forests in a graph can be very large (especially for graphs generated from datasets with thousands of genes) this option is excluded for practical purposes on large datasets.

In this preliminary application, the computation of the Shapley value on the graph-restricted game may be done by means of visual inspection of the graph, looking at the shortest paths (depicted in green in Figure 3) which connect nodes of unanimity coalitions (yellow nodes in Figure 3). It is in fact easy to check that component $\{1, 3\}$ may be connected to nodes 2 and 4 only via paths which contain node 24 (in red), and node 2 may be connected to node 4 only via paths which contain node 10 or node 22. Consequently, by relation (7), the graph-restricted game is

$$w_\Gamma^v = u_{\{1,2,3,4,10,24\}} + u_{\{1,2,3,4,22,24\}} - u_{\{1,2,3,4,10,22,24\}} + u_{\{26,27,28\}} + u_{\{38\}} + u_{\{72\}}$$

and, by relation (2), $\phi_1(w_\Gamma^v) = \phi_2(w_\Gamma^v) = \phi_3(w_\Gamma^v) = \phi_4(w_\Gamma^v) = \phi_{24}(w_\Gamma^v) = \frac{2}{6} - \frac{1}{7} = \frac{4}{21}$, $\phi_{10}(w_\Gamma^v) = \phi_{22}(w_\Gamma^v) = \frac{1}{6} - \frac{1}{7} = \frac{1}{42}$, $\phi_{26}(w_\Gamma^v) = \phi_{27}(w_\Gamma^v) = \phi_{28}(w_\Gamma^v) = \frac{1}{3}$, $\phi_{38}(w_\Gamma^v) = \phi_{72}(w_\Gamma^v) = 1$ and $\phi_i(w_\Gamma^v) = 0$ for each other gene i.

By relation (4) and the above calculations, only three genes have a γ centrality measure larger than zero, i.e. OR2B2, SCD and ODF4 (Table 1 and Figure 3, red nodes). We focus on genes with strictly positive γ because they represent those genes with a positive differential power between the graph-restricted game and the association game. In other words, we are interested to select those genes whose power increase as a consequence of their interactions in the network. From the values of γ centrality reported in Table 1 we argue that genes SCD and ODF4 have the same importance, whereas OR2B2 is eight times more critical than the other two in guaranteing the interconnection of the associated genes.

Such genes are connected to genes associated to the key-gene TP53. This is a consequence of the fact the other three key genes do not contribute to γ centrality, being the terms $\phi(u_{\{26,27,28\}}) + \phi(u_{\{38\}}) + \phi(u_{\{72\}})$ both in $\phi(v)$ and $\phi(w_\Gamma^v)$.

Among genes with positive γ, gene OR2B2 encodes for an olfactory receptor protein which is member of a large family of G-protein-coupled receptors. G proteins have been suggested to be involved in the respiratory burst (release of ROS) caused by asbestos (Elferink and Ebbenhout (1988)). The principal product of SCD is oleic acid, which is formed by desaturation of stearic acid. The ratio of stearic acid to oleic acid has been implicated in the regulation of cell growth and differentiation through effects on cell membrane fluidity and signal transduction. ODF4 encodes a protein that is localized in the outer dense fibers of the tails of mature sperm. As a functional annotation, all such genes encode for transmembrane proteins.

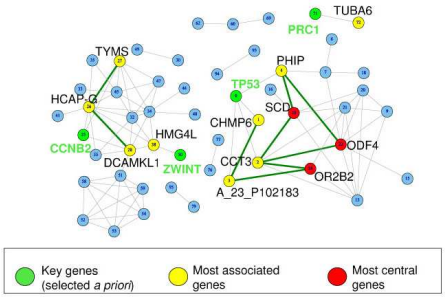


Fig. 3. Interaction network between genes (nodes). Interactions between gene's pairs are represented by edges. Isolated genes were removed. Thicker edges show the shortest paths among the most associated genes. Most central genes according to the γ measure of centrality are shown.

Table 1. Genes with γ centrality measure greater than zero. Methods were implemented using R language (R Development Core Team (2004))

ID	Symbol	Name	γ Centrality
24	OR2B2	olfactory receptor, family 2, subfamily B, member 2	$\frac{4}{21}$
10	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	$\frac{1}{42}$
22	ODF4	outer dense fiber of sperm tails 4	$\frac{1}{42}$

4 APPROXIMATE COMPUTATION

As we already observed in the last section, the implementation of algorithms aimed to generate the set of all forests in a real biological network is an unpractical approach because of the huge storage memory and the computational burden. This section is devoted to the description of an alternative approach based on approximate calculations, and to its application on a large biological network where also other centrality measures are applied.

Let (N, K, I) be a gkg situation and let $\langle N, \Gamma\rangle$ be an interaction network. For instance, with $|N| = 15$, $|K| = 1$, $|I| = 5$ and $|\Gamma| = 21$ (i.e., $\langle N, \Gamma\rangle$ has a graph density equal to 0.2), the exact computation of the Shapley value of the restricted game w_Γ^v according to relation (8) and our R language implementation, required less than two minutes (on a PC with a 2 GHz processor and 2 GB of memory). But the problem explodes exponentially in time on more dense graphs.

For this reason, in order to make feasible (and reasonable in terms of elapsed time) the application of the method also to larger biological networks, we avoid the exhaustive generation of all

forests and the consequent exact computation of the Shapley value of a restricted game. Alternatively, we limit our analysis to a smaller subset of forests, randomly selected from $\mathcal{F}(\Gamma)$ by means of a probability sampling method which assigns a uniform probability distribution to all subgraphs with a predefined number k of edges. The value k is given by the cardinality of the set of edges in a forest in $\mathcal{F}(\Gamma)$.

Given a set $\mathcal{R}(\Gamma) \subset \mathcal{F}(\Gamma)$ of forests randomly generated according to such a procedure, the set of the minimal components containing S in $\langle N, \Gamma \rangle$ is denoted by $\bar{\mathcal{M}}_\Gamma(S) = \{M_{F_\Gamma}(S) | F_\Gamma \in \mathcal{R}(\Gamma)\}$ and the approximated Shapley value of w_Γ^v is computed according to equation (8) with $\bar{\mathcal{M}}_\Gamma$ in the role of $\mathcal{M}_\Gamma$.

According to this procedure, only 100 randomly selected samples were needed to calculate the exact value of γ centrality for genes presented in the preliminary application introduced in Section 3 (elapsed time less than one second).

This random sampling method was also used to calculate an approximated γ centrality for a lager graph with 235 nodes and 2690 edges. Only one key gene (again gene TP53) was considered, on the same dataset introduced in Section 3. In this case, 300 genes with the highest absolute value of Pearson correlation with TP53 were initially selected for further analysis. From the gene expressions of the selected 300 genes, following the same method described in Section 3, a network was constructed. More precisely, a link between two nodes was established if and only if their absolute Pearson correlation coefficient was greater than 0.75. Only genes connected (directly or via other nodes) to TP53 were considered (finally, $|N| = 235$). We focused exclusively on the component connected to key gene TP53 because the contribution of the other key genes to γ centrality in a larger network (constructed according to the procedure previously described) is the same as it was calculated at the end of Section 3 on the network depicted in Figure 3, that is null.

The algorithm for the approximated computation of γ centrality was applied to the generated interaction network using 1000 random samples. In addition, in order to further simplify calculations of the Shapley value via relation (8), only minimal components with not more than 10 nodes were considered in $\bar{\mathcal{M}}_\Gamma$.

Only 42 genes showed an approximated γ centrality strictly positive (indeed, representing genes with a positive differential power). Those findings were compared with the most 42 central genes according to other four common measures of centrality. In the following, we briefly introduce those measures. In order to do that, we denote by $d(u, v)$ the minimum number of edges to connect two nodes u and v in $\langle N, \Gamma \rangle$:

1) Degree centrality (Shaw (1954); Nieminen (1974)): the degree centrality of $v \in N$ is defined as the number of edges in $e \in \Gamma$ such that $v \in e$.
2) Closeness centrality (Beauchamp (1965); Sabidussi (1966)): the closeness centrality of node $v \in N$ is defined as $\frac{|N|-1}{\sum_{y \in N} d(v,y)}$. therefore it measures the extent to which node $v \in N$ is close to all other nodes in the $\langle N, \Gamma \rangle$.
3) Betweenness centrality (Bavelas (1948); Freeman (1977)): let $u, v, z \in N$ and let $n_{u,v}$ be the number of paths formed by precisely $d(u, v)$ edges and Let $n_{u,v}(z)$ be the number of paths formed by precisely $d(u, v)$ edges which contains node z. The rate of communication between u and v that can be monitored by an interior node z is denoted by $\delta_{u,v}(z) = n_{u,v}(z)/n_{u,v}$. If no shortest path between u and v exists $\delta_{u,v}(z) = 0$ by definition. The betweenness centrality of z is defined as $\sum_{u,v \in N, u \neq v, u \neq z, v \neq z} \delta_{u,v}(z)$.
4) Eigenvector centrality (Bonacich (1972)): Let $v \in N$. Then the eigenvector centrality of v is defined as the v^{th} element of the principal eigenvector of the adjacency matrix corresponding to $\langle N, \Gamma \rangle$. This principal eigenvector is normalized such that its largest entry is 1. This centrality is a measure for how well connected a node is to other highly connected nodes in a network.

For each pair of centrality measures considered, the percentage of common genes among the first 42 with highest centrality for each measure are reported in Table 2. Note that the betweenness centrality has the maximum level of overlap with the list of genes ranked according to the approximated γ centrality. Ten genes are ranked among the first 42 with highest centrality according to all centrality measures (i.e., UBE1, STK23, ODF4, RNF170, SPAG9, TMCC2, C1QTNF9, SAMD4B, HEATR2 and LPHN1).

The most 12 central genes according to γ are reported on Table 3. Note that genes SCD, ODF4 and OR2B2, that were founded in the analysis of the smaller network introduced in Section 3, are still ranked among the most central genes, and other centrality measures support this result (see Table 3). Among genes that are predicted to be central only by gamma centrality (namely, MPP1, SIAH2, EDIL3, OR1K1, TAS1R2), we observe that the human homolog of Drosophila SIAH2 has been studied in literature in connection with anticancer agents (Atique *et al*. (2008)), as well as SIAH1 has been related to p53-mediated apoptosis (Relaix *et al*. (2000)); EDIL3 acts as an angiogenic factor in the context of solid tumor formation (Aoka *et al*. (2002)).

In addition, from Table 3 it seems that γ centrality behaves very close to betweenness centrality, at least with respect to the most central genes (six genes in common among the top ten). In order to understand whether this behavior is affected by the random selection of edges in the procedure for γ centrality approximation, we assessed the impact of the number of random samples used in the procedure described in Section 4 on the overlap with the most central genes from the different centrality measures. Specifically, for different numbers of random samples m, with $m = 200, 400, 600, 800, 1000, 1200$, we applied our algorithm for γ centrality approximation to the same network. For each m, the ten genes with highest approximated γ centrality were selected and compared with the ten most central genes according to the other centrality measures. The comparison of top ten lists was reiterated ten times, for each m. Results are summarized in Figure 4. Note that already after 400 random samples, the overlap ratios seem to stabilize and, for each value m, the list provided by the approximated γ centrality shows the highest overlaps with the lists provided by betweenness centrality and closeness centrality.

In some instances, we also observed that the sensibility of the method based on γ centrality to structural variation of the network is comparable to that of the other centrality methods analyzed in this study. For example, if a new component connected to gene TP53 is considered, with 240 nodes and 2888 edges (defined on a Pearson correlation cutoff of 0.745) only three genes out of the 12 most central ones according to γ were detected also among the

Table 2. Percentage of common findings among lists of 42 genes with highest centrality according to different centrality measures. The average overlap of each measure with the others is shown.

	appr. γ	Deg.	Clos.	Bet.	Eigen.
approxim. γ	*	43%	48%	48%	45%
Deg.	43%	*	83%	31%	95%
Clos.	48%	83%	*	38%	81%
Bet.	48%	31%	38%	*	31%
Eigen.	45%	95%	81%	31%	*
average	46%	63%%	61%	37%	64%

Table 3. Most 12 central genes according to γ centrality. Numbers shows genes found among the 12 most central genes according to degree centrality ([1]), closeness centrality ([2]), betweenness centrality ([3]), eigenvector centrality ([4]).

Symbol	Name	appr. γ
UBE1 [1,2,3]	ubiquitin-activating enzyme E1	0.0309
STK23 [1,2,3]	serine/threonine kinase 23	0.0119
MPP1	membrane protein, palmitoylated 1, 55kDa	0.0072
SCD [3]	stearoyl-CoA desaturase (delta-9-desaturase)	0.0047
ODF4 [1,2,3]	outer dense fiber of sperm tails 4	0.0046
RNF170 [2,3]	ring finger protein 170	0.0038
PLXNA3 [1,4]	plexin A3	0.0037
SIAH2	seven in absentia homolog 2 (Drosophila)	0.0035
OR2B2 [3]	olfactory receptor, family 2, subfam. B, memb. 2	0.0033
EDIL3	EGF-like repeats and discoidin I-like domains 3	0.0032
OR1K1	olfactory receptor, family 1, subfam. K, memb. 1	0.0032
TAS1R2	taste receptor, type 1, member 2	0.0032

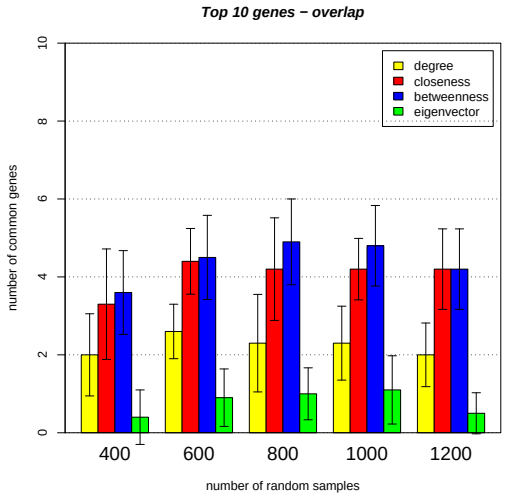
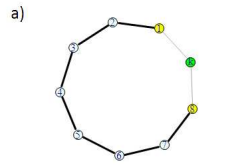
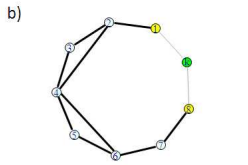


Fig. 4. Common findings among lists of 10 genes with highest centrality according to different centrality measures. The average number of common genes (and the respective standard deviations) of γ centrality with each other centrality measure computed over ten iterations of the approximation algorithm are shown, for different numbers of randomly selected samples.



gene	γ	Deg.	Clos.	Bet.	Eigen.
1	-0.38	1	0.25	0	0.35
2	0.13	2	0.32	6	0.65
3	0.13	2	0.39	10	0.88
4	0.13	2	0.44	12	1.00
5	0.13	2	0.44	12	1.00
6	0.13	2	0.39	10	0.88
7	0.13	2	0.32	6	0.65
8	-0.38	1	0.25	0	0.35



gene	γ	Deg.	Clos.	Bet.	Eigen.
1	-0.33	1	0.35	0	0.26
2	0.17	3	0.50	6	0.70
3	0.00	2	0.47	0	0.63
4	0.17	4	0.64	12	1.00
5	0.00	2	0.50	0	0.64
6	0.17	3	0.58	10	0.72
7	0.17	2	0.44	6	0.31
8	-0.33	1	0.32	0	0.12

Fig. 5. Two different interaction networks (a and b) with eight genes (interactions are represented by thick lines). Gene 1 and 8 are the most associated genes in both networks, which directly interact (thin lines) to the key-gene k. Centrality values of nodes according to different centrality measures are shown in the corresponding tables.

most central ones on the original network (reported in Table 2). In comparison, the number of genes that were found in both lists when the other centrality measures were applied to both networks was 2 for degree centrality, 1 for closeness centrality, 3 for betwenness centrality and 2 for eigenvector centrality.

From Table 2 it turns out that the list of most central genes according to the approximated γ centrality shares a similar number of genes with all the lists of genes generated by the other centrality measures (between 43% and 48%). This result is not surprising, but it can be explained by means of the very basic properties of the γ index. The definition of γ centrality, based on the notions of minimal components for coalitions (see relation 8), generalizes the idea of number of paired relations (degree centrality), all shortest paths (closeness centrality) or geodesic paths (betweenness centrality). The interaction networks depicted in Figure 5 clarify this point. In the interaction network of Figure 5.a, where all genes between 2 and 7 are needed to connect the associated genes 1 and 8, it is not natural to discard the possibility of using longer paths, simply because shorter ones exist. Then, on this network, γ centrality behaves similar to degree centrality, providing the same level of importance to all genes between 2 and 7 (differently from the other measures, which assign the biggest amount of importance to nodes 4 and 5 and the smallest amount to 2 and 7). On the other hand, we would tend to discard a long path between two genes, in favor of a one-edge path, because in this case it imposes additional intermediaries genes which are not needed to connect associated genes. This is the case of the interaction network depicted in Figure 5.b, where genes 3 and 5 are intermediary genes not necessary to connect associated genes 1 and 8, and therefore they receive a null level of centrality both from γ and betweenness centralities, whereas the other measures give an intermediate level of centralities to such nodes.

5 CONCLUSION

In this paper, a new measure of the importance of genes in biological networks based on coalitional games is introduced. The new measure, calculated from the Shapley value of two coalitional games, has been used to express the centrality of each gene in interaction with the others and keeping into account *a priori* knowledge about genes playing a key function on a certain biological process.

An approximation method for the calculation of γ centrality in practical biological networks is also presented. According to this procedure, the generation of all spanning forests in a biological network is not needed, but the analysis is limited to a smaller subset randomly selected from the set of all forests. Results provided by the approximation procedure on biological networks are discussed and compared to the results provided by the application of classical centrality measures.

The use of γ index as a centrality measure is supported by the basic intuition that it is a difference of power indices between a situation where binary interactions are considered (i.e., the graph-restriction game) and another one where they are not (i.e., the association game), and by the comparison with properties related to other centrality measures on some examples. In order to generalize these argumentations, an important issue for future research is to address a comprehensive analysis of the properties satisfied by the Shapley value on graph-restricted games, with the objective to better contextualize its interpretation as a centrality measure.

6 ACKNOWLEDGEMENTS

The authors are grateful to three anonymous referees for their helpful comments.

REFERENCES

Albino,D., *et al*. (2008) Identification of low intratumoral gene expression heterogeneity in Neuroblastic Tumors by wide-genome expression analysis and game theory. *Cancer*, **113**, 1412-22.

Amaratunga,D., and Cabrera,J. (2004). Exploration and Analysis of DNA Microarray and Protein Array Data, Wiley-Interscience, New Jersey.

Aoka,Y., *et al*. (2002) The embryonic angiogenic factor Del1 accelerates tumor growth by enhancing vascular formation. *Microvascular Research*, **64**, 148-61.

Atique,U., *et al*. (2008) Effect of Disrupting Seven-in-Absentia Homolog 2 Function on Lung Cancer Cell Growth. *Journal of the National Cancer Institute*, **100**, 1606-29.

Bavelas,A. (1948) A mathematical model for small group structures. *Human Organization*, **7**, 16-30.

Beauchamp,M. (1965) An improved index of centrality. *Behavioral Science*, **10**, 161-163.

Bonacich,P. (1972) Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, **2**, 113-120.

Carlson,M., *et al*. (2006) Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics*, **7**, 40.

Elferink,J.G., and Ebbenhout,J.L. (1988) Asbestos-induced activation of the respiratory burst in rabbit neutrophils. *Research Communications in Chemical Pathology & Pharmacology*, **61**, 201-211.

Fragnelli,V., and Moretti,S. (2008) A game theoretical approach to the classification problem in gene expression data analysis, *Computers & Mathematics with Applications*, **55**, 950-959.

Freeman,L.C. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**, 3541.

Gabow,H.N., and Myers,E.W. (1978) Finding all spanning trees of directed and undirected graphs. *SIAM Journal on Computing*, **7**, 280-287.

Golub,T., *et al*. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.

Gòmez,D. *et al*. (2003) Centrality and Power in Social Networks: a Game Theoretic Approach. *Mathematical Social Sciences*, **46**, 27-54.

Gòmez,D. G*et al*. (2004) Splitting graphs when calculating Myerson value for pure overhead games. *Mathematical Methods of Operations Research*, **59**, 479-489.

Jeong,H. *et al*. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41-42.

Junker,B.H., *et al*. (2006) Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*, **7**, 219.

Kapoor,S., and Ramesh,H. (1995) Algorithms for enumerating all spanning trees of undirected and weighted graphs. *SIAM Journal on Computing*, **24**, 247-265.

van Leeuwen,D., *et al*. (2008) Genomic analysis suggests higher susceptibility of children to air pollution. *Carcinogenesis*, **29**, 977 - 983.

Lucchetti,R., *et al*. (2009) The Shapley and Banzhaf indices in microarray games. *Computers & Operations Research*, **37**, 1406-1412.

Mason,O., and Verwoerd,M. (2007) Graph theory and networks in biology. *IET Systems Biology*, **12**, 89-119.

Minty,G.J. (1965). A simple algorithm for listing all the trees of a graph. *IEEE Transactions on Circuit Theory*, **CT-12**, 120.

Moretti,S. (2009) Game Theory applied to gene expression analysis. *4OR: A Quarterly Journal of Operations Research*, **7**, 195-198.

Moretti,S. (2010) Statistical analysis of the Shapley value for microarray games *Computers & Operations Research*, **37**, 1413-1418.

Moretti,S., *et al*. (2008) Combining Shapley value and statistics to the analysis of gene expression data in children exposed to air pollution. *BMC Bioinformatics*, **9**, 361.

Moretti,S., and Patrone,F. (2008) Transversality of the Shapley value. *Top*, **16**, 1-41

Moretti,S., *et al*. (2007) The class of microarray games and the relevance index for genes, Top, **15**, 256-280.

Myerson,R.B. (1977) Graphs and cooperation in games. *Mathematics of Operation Research*, **2**, 225-229.

Nieminen,J. (1974) On centrality in a graph. *Scandinavian Journal of Psychology*, **15**, 322-336.

Owen,G. (1995) Game Theory. 3rd edn. Academic Press.

Parmigiani,G. *et al*. (2003). The analysis of gene expression data: an overview of Methods and Software. *The analysis of gene expression data: methods and software*, G. Parmigiani *et al*. edn., New York.

R Development Core Team (2004). R: A language and environment for statistical, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Relaix,F., *et al*. (2000) Pw1/Peg3 is a potential cell death mediator and cooperates with Siah1a in p53-mediated apoptosis. *Proceedings of the National Academy of Sciences*, **97**, 2105-2110.

Sabidussi,G. (1966) The centrality index of a graph. *Psychometrika*, **31**, 581-603.

Shapley,L.S. (1953). A Value for n-Person Games. *Contributions to the Theory of Games II* (Annals of Mathematics Studies 28), H. W. Kuhn and A. W. Tucker edn., Princeton University Press, pp. 307-317.

Shapley,L.S., and Shubik,M. (1954) A Method for Evaluating the Distribution of Power in a Committee System. *American Political Science Review*, **48**, 787-792.

Shaw,M.E. (1954) Grou*p strucure and the behaviour of individuals in small groups.* Journal of Psychology, **38**, 139-149.

Storey,J.D., and Tibshirani,R. (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarray. *The Analysis Of Gene Expression Data: Methods And Software*. G. Parmigiani *et al*. edn. Springer, New York.

Tusher,V.G., *et al*. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, **98**, 5116-5121.

Zhang,B., and Horvath,S. (2005) A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, **4**, Article 17.