

# Role and position detection in networks: reloaded

Davide Vega, Roc Meseguer and Felix Freitag  
Universitat Politcnica de Catalunya  
{dvega,meseguer,felix}@ac.upc.edu

Matteo Magnani  
Uppsala University  
matteo.magnani@it.uu.se

**Abstract**—Roles and positions are structural components in complex social systems which group actors based on how similarly they are connected to the rest of the actors. Role and position detection methods have been successfully used to evaluate and understand the dynamics of social networks and the behavior of their members. However, actor similarities used to detect positions have been based on pairwise comparisons so far: e.g., *structural equivalence* states that Alice and Bob are in the same position if they are both connected or not to the same other actors in the network, one by one. In this work we present a new framework to find positions and roles using *comparisons between actors and sets of actors* instead of just using pairwise comparisons. In this way we enable the usage of many more measures of similarity inside position and role detection methods, e.g., based on distances, community structure, triangles and cliques. As a result, we can identify new types of easily interpretable positions. Additionally, the proposed idea can be adapted to more complex models like hypergraphs or multiplex/multi-relational networks. We have evaluated our work on both synthetic and real data, using several existing and new similarity measures and providing both qualitative and quantitative evidence of the new possibilities enabled by our approach.

## I. INTRODUCTION

Structural analysis on networks intends to capture and interpret how nodes are related according to the network topology. When applied to social networks, structural analysis is able to identify key actors<sup>1</sup> or groups of actors whose connectivity influences somehow the dynamics of the system. Three typical ways of grouping actors based on their connections consist in identifying *communities*, *positions* and *roles*. While related, these are three distinct types of groups and they typically require distinct algorithmic treatments.

To briefly recall the difference between community, position and role, we use Padgett's social network representing business relationships among Florentine families during Renaissance [1] as an example (see Fig. 1). A *community* indicates a cohesive group of actors, with many connections inside the group and fewer relationships with other actors outside it. As an example, the five nodes on the top of the figure form a community. *Roles and positions*, instead, focus on the interchangeability of the actors, and do not require any internal connectivity. In its simplest form, a *position* is defined as a group of actors who are similarly connected to other actors in the network. In our example, the families *Salvati*, *Pazzi* and *Tornabuon* are in the same position, because they are all connected to the *Medici* family and to no other family. *Roles*, instead, refer to actors with similar patterns of connectivity, independently of the specific actors to whom they are connected. In our running example, the *Barbadori* family

has the role of connecting two otherwise disjoint parts of the network. From this point of view, it does not matter who exactly is connected to them: if the *Barbadori* family were connected to *Salviati* instead of *Ginori*, it would still play the *same role* in the network, but *from a different position*.

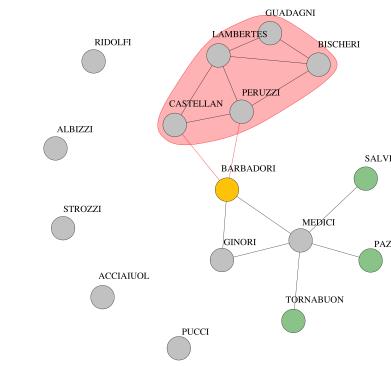


Fig. 1: Padgett's business family network, where we have highlighted a community (red area), a structurally equivalent position (green nodes), and a bridging role (yellow node)

In this work we focus on position and role analysis. To understand our contribution we can consider a standard mathematical definition of the concept of position. In the model based on the so-called *structural equivalence*, two actors are in the same position if and only if they are connected with the exact same actors. Let  $G = (U, E)$  be a graph representing a social network, where  $U$  is a set of nodes representing actors and  $E(i, j) = 1$  if nodes  $i$  and  $j$  are connected, 0 otherwise. Then, we can say that two nodes  $i$  and  $j$  are structurally equivalent (and so in the same position) if and only if [2]:

$$E(i, k) = E(j, k) \quad \forall k \in U; k \neq i, j \quad (1)$$

In our example, *Salvati* and *Pazzi* are both connected to *Medici* (so,  $E(\text{Salvati}, \text{Medici}) = E(\text{Pazzi}, \text{Medici}) = 1$ ), and for every other actor  $k$  they are not connected to it, so  $E$  equals 0 for both families in all other cases.

This basic definition has been extended in many ways in the literature [3], e.g., replacing  $E$  with other comparison measures (that we generically denote as  $D$  in the following) or using this formula in an iterative process, as reported in Section IV. However all existing variations rely on comparing a pair of actors against *one single item at a time*, here represented by the symbol  $k$ . In this paper we extend this perspective on position

<sup>1</sup>Individuals or organizations corresponding to the nodes in the social graph.

analysis by replacing Eq. 1 (and its generalizations) with a set-based definition. Two nodes  $i$  and  $j$  are, according to our model, in the same position if and only if:

$$D(i, S_k) = D(j, S_k) \quad \forall S_k \subseteq U \quad (2)$$

The main formal difference is the usage of a set  $S_k$  instead of the singleton  $k$ : to decide if  $i$  and  $j$  belong to the same position, we check how they are related not just to other single actors, but to groups of them. This allows us to use several more measures ( $D$ ) to compare actors, that would not make sense for single pairs of nodes (like  $E(i, j)$ ) and have thus been overlooked in the literature on position detection.

As a concrete example, consider Fig. 2, showing Padgett's marriage network with each family colored according to its social position. Families *Albizzi* and *Guadagni* are connected to totally different nodes, that are themselves in different positions. So, they would not be considered being part of the same position by existing methods. For example, if we check the *Medici* family, *Albizzi* is connected to it while *Guadagni* is not. If we check *Lambertes*, *Albizzi* is not connected to it while *Guadagni* is. However, if we now consider the pair  $\{Lambertes, Medici\}$  and a comparison function:

$$D(i, \{k, q\}) = \begin{cases} 1 & \text{if } i \in \text{short. path betw. k and q} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

we can see that both *Albizzi* and *Guadagni* are on a shortest path between them. If we check other pairs of nodes, we can see that this is true in several other cases, e.g., to efficiently go from *Bischeri* to *Ginori* we should also pass through *Albizzi* and *Guadagni*. In summary, *Albizzi* and *Guadagni* are included in the same position because they share the same relationship with other *pairs of nodes*, instead of single individuals.

This example highlights three important aspects of our work. First, using this set-based comparison and the corresponding distance-based function we are identifying a **new type of position**. Second, this extension allows us to use several **other network measures** involving comparisons with larger sets of actors, like triangles, cliques and communities – we will introduce all these additional possibilities in the remainder of the paper. Third, we obtain a cleaner understanding of the often ambiguous **relationship between positions and roles**: with this extended definition, we can say that for each type of position we have a corresponding role, which considers the pattern of connectivity but not the specific individuals involved. In our example, the *Ridolfi* and *Strozzi* families might also constitute a position, different from the previous one – e.g., they are in the shortest path between *Peruzzi* and *Medici*, which is not the case for *Albizzi*. However, the number of the shortest paths traversing them (that is, their betweenness) is similar, making them play a similar role in the network. It is worth noticing that to the best of our knowledge concepts like betweenness had never been related to the concept of position before, while this connection directly follows from our extended formalization once Eq. 3 is used as a comparison function.

As a future-looking note, our extended model also gives us the flexibility to study social roles and positions in other network models like **hypergraphs**, which can be represented

as hyperlink adjacency matrices without losing information about the hyperlinks – as it would happen if the regular adjacency matrix is used. Another example are **multiplex/multi-relational** graphs, where our model enables the usage of measures based on paths traversing multiple types of relational ties.

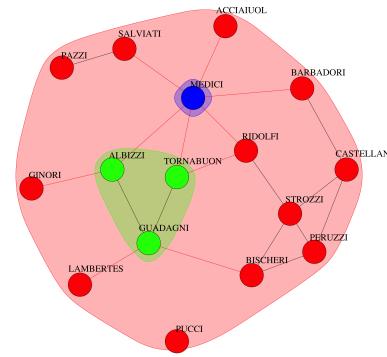


Fig. 2: Padgett's marriage family network and approximate positions defined as being part of the shortest path connecting pairs of nodes

In summary, the main contributions of this paper can be summarized as follows.

- We introduce a novel framework to identify new types of social positions and roles based on well-known network measures (Section II).
- We validate our model both qualitatively and quantitatively, using real and synthetic data (Section III).

In our quantitative validation, we compared the positions found in a real network with the ones found in synthetic networks generated according to different models. The collected evidence shows that our approaches do not just identify randomly occurring behaviors, but positions that are only found in real social networks (Section III-C).

## II. A NEW BLOCKMODELING FRAMEWORK

Blockmodeling [4] is, to our knowledge, the most used and explored technique to detect roles and positions in social networks and, more generally, in any system that can be modeled mathematically using a graph. In blockmodeling, actors are grouped into *positions* – called blocks, sometimes roles – based on a similarity or dissimilarity measure between them. To compute this measure, actors are compared based on their connectivity in the network. In its original form, the similarity measure corresponds to the correlation between columns in the graph adjacency matrix, which results in including actors connected to the same other actors into the same position.

In this section we describe our *framework for group relations* which allows us to apply blockmodeling analysis to find social roles and positions based on the global structure of the network, rather than being constrained by pairwise comparisons. The framework has two basic components: the

*extended comparison function* for group measures and the *computing algorithm* for identifying positions and roles.

The *extended comparison function* is a two-dimensional matrix ( $M$ ) that stores the similarity or dissimilarity between actors (rows) and sets of actors (columns). The *computing algorithm* used in our experiments is a generalization of the REGE/A algorithm proposed in [5] for regular equivalence. In general, any clustering algorithm already used in unsupervised blockmodeling analysis could be used instead.

#### A. Extended comparison function

In order to compare actors with subsets of actors we replaced the adjacency matrix with a more complex structure, capable of storing extended relations. It will be used as input data for the computing algorithm. Building it involves two interdependent steps: a) dividing the actors into groups of interest and b) defining the comparison measure.

We have mentioned how our approach is based on a generic comparison function:

$$D : (U, S) \rightarrow \mathbb{R} \quad (4)$$

defined over a graph  $G = (U, E)$ , where  $S \subseteq 2^U$  depends on  $D$ . As in Eq. 3,  $S$  consists of all subsets of  $U$  of cardinality 2. We can then use the function  $D$  to build our extended matrix  $M$ , by substituting the adjacency matrix equivalence in Eq. 1 with our formula. Hence, we can define the extended matrix  $M$  as:

$$M(i, S_j) = D(i, S_j) \quad (5)$$

Back to our example on detecting positions as *being part of the shortest path that connects pairs of nodes*, Fig. 3 shows the corresponding extended matrix  $M$ . Each cell in the matrix corresponds to one binary relation between an actor  $i$  and a set of two other actors  $(k, q)$ .

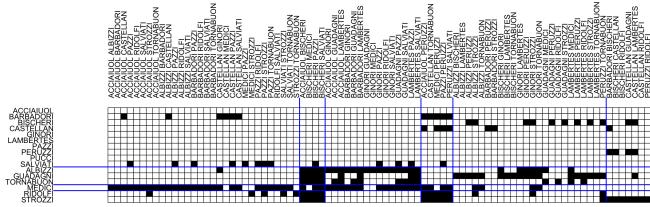


Fig. 3: Extended relations matrix for the Padgett's marriage family network.

#### B. Positions and role assignment

After computing our extended comparison matrix, which associates each actor in the network with the subsets of actors used to measure the equivalence, the next step consists in partitioning the actors into  $\beta_M = \beta_1, \beta_2, \dots, \beta_m$  similar positions and  $\rho_Z = \rho_1, \rho_2, \rho_3, \dots, \rho_z$  similar roles.

Positions are computed by clustering the rows of the extended matrix and forming groups of actors whose relation with the same subsets are similar. In the literature there are

many clustering algorithms that can be used to make this assignment. In order to simplify the results, and for comparison purposes with other indirect blockmodeling methods, we decided to use hierarchical clustering. In concrete, we first generate a dissimilarity matrix by computing the euclidean distance between rows of our extended matrix  $M$  and then we generate a hierarchical clustering using the Ward [6] cluster similarity function. Finding the optimal number of clusters (i.e., positions) is discussed in detail in Section II-C.

The procedure to identify roles uses the same extended matrix  $M$  as input to compute actors' patterns of relations – that is, some summary of the distribution of row values. Then, actors are grouped using any clustering algorithm. In case of using a binary extended matrix, for example, it is enough to count the number of 0s and 1s and compare them.

TABLE I: Roles identified as being part of the shortest path between pairs of nodes in Padgett's marriage families network.

$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$	$\rho_5$
ACCIAIUOL	CASTELLAN	BARBADORI	ALBIZZI	MEDICI
GINORI	PERUZZI	BISCHERI	GUADAGNI	
LAMBERTES		SALVIATI	TORNABUON	
PAZZI			RIDOLFI	
PUCCI			STROZZI	
0	4, 7	11 - 17	26, 27	50

As an example of positional analysis, we used our method and depicted in Fig. 3 the resulting assignment of actors that are *being part of the shortest path that connects pairs of nodes* into four positions. Table I, instead, shows the corresponding roles detected using the same equivalence. Notice that the two assignments are not identical. While the *Medici* family, as an example, is the single member of a role and a position, *Strozzi* and *Tornabuon*, who play the same social role, are in fact in different positions – because they are in the same number of the shortest paths, but between different sets of actors. Under a strict check of equivalences between the rows of the extended matrix, positions would be finer partitions of the roles.

#### C. Approximate positions and roles

As we mentioned, each definition of equivalence admits different degrees of freedom, because in most cases – especially when the notion of structural equivalence is used – the normal variability in connectivity prevents us from finding two nodes with many connections and connected with the exact same other nodes. To relax the definition, and find meaningful positions and roles using indirect approaches, it is sometimes necessary to utilize some knowledge about the social network under analysis, e.g., specifying the number of expected positions.

As an alternative, we can obtain different numbers of positions by cutting the dendrogram generated by the clustering algorithm (see Fig. 4) at different heights, indicated as  $h$  in the following. Cutting the dendrogram at height  $h = 0\%$  we will find non-approximated positions, e.g., two actors must be indistinguishable according to the  $D$  function to be included in the same position. As we pull up the cutting point we can group actors into fewer positions, where more and more differences in the values of  $D$  are allowed among actors in the same position.



Fig. 4: Dendrogram representing the hierarchical clustering of the rows of the matrix in Fig. 3 based on actor similarity

### III. EXPERIMENTAL ANALYSIS

In order to evaluate our methodology we built a library in R using the *blockmodeling-package* [7] as a baseline. The library has been used to perform the experimental analysis, which included the detection of roles and positions with different combinations of graphs and equivalences. In this section we first describe the networks used for the evaluation and the comparison of positions detected on both social and synthetic networks. Finally, we qualitatively compare the results of the different extended measures proposed.

#### A. Datasets

We evaluated our proposal using a set of real social networks – representative of the social structures that arise and are well known in the literature. Additionally, we created a set of synthetic networks for comparison purposes.

**Florentine families** [1]: contains a two-relational graph describing the social relations among Renaissance florentine families (person aggregates) collected by John Padgett from historical documents.

**AUCS** [8]: contains a five-relational graph describing the social relations among employees of a Computer Science department. For the analysis we have flattened the network into a mono-relational graph.

**Synthetic data**: contains a set of networks built using the Erdos-Renyi [9] and Barabasi [10] models, with the same number of nodes as our real networks, but varying densities.

#### B. Other extended relations as equivalences

We have previously described our framework using the example measure of **being in the shortest path (BSP)**, which intends to capture actors between common pairs of actors. Next we describe a set of easily interpretable measures, to demonstrate the possibilities of our framework, and how to build their corresponding matrix  $M$ .

**Clique connectivity (CC) and Minimum Clique connectivity (MCC)** intend to identify groups of actors with the same

ties to the same cliques. Both measures use cliques as the sets of actors required by our method, but differ in how to compute the equivalence: the former computes the number of ties between each actor  $u$  and a given clique, while the latter only considers binary relations: 1 if  $u$  has at least one tie with some member of the clique, 0 otherwise.

**Community connectivity (COMC) and Others' community connectivity (OCOMC)** intend to identify groups of actors with the same ties to the same communities. Hence, the matrix  $M$  contains the communities as columns, and the percentage of ties to their members as values. However, while COMC considers all communities, OCOMC discards ties to the community to which the node under evaluation belongs. In the following experiments communities are computed maximizing modularity.

#### C. Framework validation

In the following we investigate whether the roles and positions found in real social networks and synthetic graphs differ. If they did not, then we could question the ability of our approach to identify meaningful social structures. Synthetic graphs have been built according to the Erdos-Renyi (ER-p) and Barabasi-Albert (BA-p) models with  $p$  denoting the probability and the exponent of the model, respectively. Each experiment has been repeated 10 times to avoid random effects due to the network formation. As the results are similar for the tested social networks, we only report those for the AUCS data.

The experiment identified three different patterns of positions. Fig. 5 shows, for each category, the number of positions found as y-axis and the corresponding height  $h$  where the dendrogram is cut (x-axis). We can observe that in all three patterns the real social network always shows positions containing multiple nodes requiring equal (Fig. 5a) or less (Figs. 5b and 5c) approximation than the synthetic graphs, and hence, closer to the ideal clustering of actors based on the definition of similarity. In general, random graphs like the Erdos-Renyi networks require more approximation when our method is used. On the other hand, BA networks are formed by a preferential attachment process, which is believed to be more similar to the interaction processes that form social networks, and hence is more likely that the positions found are capturing some social properties from the graph. However, as the number of nodes in the tested graph decreases it is more likely that structural differences become less evident.

#### D. Similarity measures analysis

The different sizes of positions depicted in Fig. 5a are explained by the flexibility, in terms of measure, of each equivalence. More sub-settings of actors will increase the number of possible row combinations, and hence, more heterogeneous positions; while fewer sub-settings will tend to create fewer positions with fewer uncertainty. On the other hand, binary comparison measures will generate more similar measures than valued ones, but probably more meaningful as they tend to cluster more actors into fewer positions.

Within the indirect blockmodeling methodology we also need to make a distinction between positions actively identified by the similarity measure and those actors that are not captured by it, e.g., disconnected nodes whose value computed by the

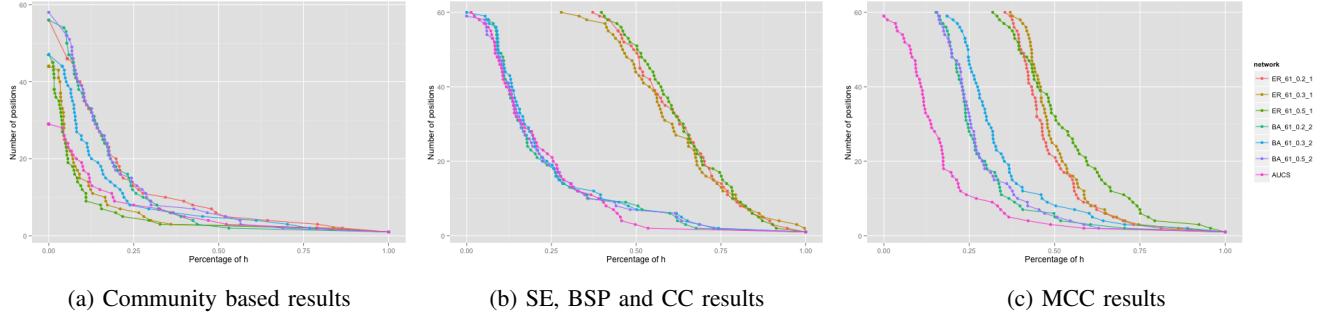


Fig. 5: Number of positions found at different heights  $h$  of the dendrogram for three extended measures. Each line represents the AUCCS network and one of the 6 synthetic graphs.

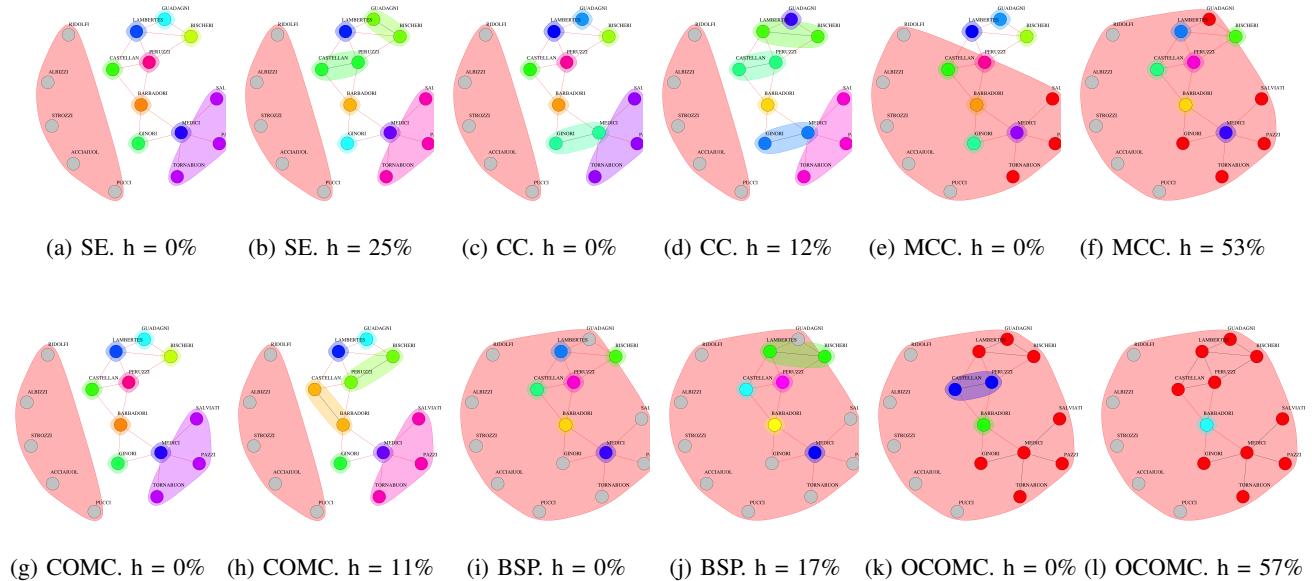


Fig. 6: Positions detected using a traditional measure – structural equivalence (SE) – and the extended measures introduced in this paper for the Business social network of Florentine families, at varying levels of approximation

measure would always be 0 and that would thus be included in a common fictitious position. As an example, in Fig. 6 6 nodes grouped together with the same color represent common positions, and we used the gray color to mark those nodes connected in such a way that is not captured by the measure at hand. We can observe that, by definition, our “being in the shortest path” extended measure places in the same position actors that are not in any shortest path between others or are completely disconnected to other actors.

The individual analysis of the measures shows that, for the used social network, the framework is able to find some non-trivial positions with no approximation using some of the tested similarity functions. As an example, the OCOMC measure finds two positions clearly containing actors connecting the upper community with the bottom one; however, the BSP measure puts each actor into a separate position. To avoid this problem, it can be useful to allow some approximation ( $h > 0$ ), and the experiments suggest that different measures may require more or less flexibility. As a general rule, higher

differences between  $h$  values are an indication that the positions found will probably be less significant.

#### IV. RELATED WORK

Both concepts of role and position have been redefined many times in the literature, by mathematicians and sociologists, more or less formally. In our work we use the definition provided by Wasserman and Faust [4]: “*In social network analysis position refers to a collection of individuals who are similarly embedded in networks of relations, while role refers to the pattern of relations which obtain between actors or between positions. The notion of position thus refers to a collection of actors who are similar in social activity, ties, or interactions, with respect to actors in other positions.*”

Despite the lack of mathematical notation, this definition clearly states the idea of identifying positions as a clustering problem where actors – vertices of a graph – are assigned to smaller subsets – called positions – based on a notion of

similarity. The similarity not only measures how similar the local connectivity between pairs of actors is in the graph, but can also measure other properties or relations.

Structural equivalence is the most basic and strict notion of similarity. Other similarities have been developed lately to relax the notion of equivalence. In *regular equivalence* [4], for example, two actors are in the same position if they have similar relations with other positions; while in *stochastic equivalence* [11] two actors are in the same position if they have the same probability distribution of ties with other actors, which is more similar to our notion of role. In a general form, we can generalize these notions of *equivalence* as a set of *node-based* features:  $A_h(i) \forall h \in H$  and *distance-based* features:  $D_p(i, j) \forall p \in P; k \neq i, j$ ; where  $H$  is a set of node attributes and  $P$  is a set of comparison functions. Notice that these relations still constrain the model to a) the adjacency connectivity matrix and b) pairwise actor comparisons  $(i, j)$ .

Several generalizations have been developed to find positions without perfect similarity/dissimilarity [11], to be used in weighted graphs [7], or even to find non-trivial equivalent positions [12]. While these approaches have proved useful to detect some kinds of positions, and are flexible enough to accommodate different kinds of similarity functions, they are also based on pairwise relationships. However, the general idea of finding approximate equivalences is also fundamental in our framework, because a strict check for equivalence would rarely identify any groups of similar actors in real social networks.

In [13] the authors describe a new taxonomy for role discovery methods, which also introduces the idea of “feature-based role discovery”. According to their proposal, the similarity between nodes can be measured using a set of node-structural features (e.g. degree, distance, etc.), which can be any set of measures taken from the initial graph. It is possible to argue that some of the extended equivalences proposed in our work could be used as features in their model, but in our framework we keep track of the relation between the measure – or feature – and the nodes related to it – the subsets of nodes that are needed to compute the measure. Because of this, our framework is able to measure not only patterns of relations (roles), but also positions.

## V. CONCLUSION AND DISCUSSION

Blockmodeling has been primarily used as a way to detect roles and/or positions in social structures using node-based measures. A lot of extensions have been proposed for blockmodeling, many of which based on replacing the original comparison function with alternative ways of measuring the network structure around the nodes. Motivated by the will of applying this traditional approach to more complex network models like multi-relational networks, we have proposed a conceptual extension of blockmodeling that allows us to plug in additional comparison functions not usable in a standard setting. At the same time, this extension enables the discovery of new kinds of positions also on simple networks: in this paper we have focused on this aspect, providing an analysis of the new possibilities and limits of our proposal.

More in detail, according to our proposal, to enable the usage of additional types of similarity functions it is necessary to extend the regular similarity matrix into a more complex

structure able to relate actors in a network with a) the extended measure and b) the extra information used to compute such measure – in this case, subsets of actors. These new measures generate a new asymmetric equivalence matrix, that can be analyzed to find both social roles and positions.

The selection of the extended measures depends entirely on the objective of the analysis and/or the meaning of the positions and roles desired. Theoretically, any measure computed in a graph as a result to apply a given function over a node and a set of nodes would be a candidate. In practice, we have observed that some measures require higher values of approximation to identify positions containing multiple nodes. In the future we plan to incorporate direct blockmodeling algorithms into our framework. Additionally, we want to extend the current proposal to other graph models, in order to understand more complex systems like multi-relational social networks. This next step will require the consideration of multi-relational measures.

## ACKNOWLEDGMENT

This work was partially supported by the European Community through the projects CONFINE: FP7-288535, and Clommunity: FP7-317879, and also by Spanish government under contract TIN2013-47245-C2-1-R, and by the Italian Ministry of Education, Universities and Research FIRB grant RBFR107725.

## REFERENCES

- [1] R. L. Breiger and P. E. Pattison, “Cumulated social roles: The duality of persons and their algebras,” *Social Networks*, vol. 8, no. 3, pp. 215 – 256, 1986.
- [2] D. R. White and K. P. Reitz, “Graph and semigroup homomorphisms on networks of relations,” *Social Networks*, vol. 5, no. 2, pp. 193 – 234, 1983.
- [3] P. Doreian, “Equivalence in a social network,” *Journal of mathematical sociology*, vol. 13, no. 3, pp. 243–281, 1988.
- [4] S. Wasserman, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.
- [5] S. P. Borgatti and M. G. Everett, “Two algorithms for computing regular equivalence,” *Social Networks*, vol. 15, no. 4, pp. 361 – 376, 1993.
- [6] F. Murtagh and P. Legendre, “Wards hierarchical agglomerative clustering method: Which algorithms implement wards criterion?” *Journal of Classification*, vol. 31, no. 3, pp. 274–295, 2014.
- [7] A. Žiberna, “Generalized blockmodeling of valued networks,” *Social networks*, vol. 29, no. 1, pp. 105–126, 2007.
- [8] L. Rossi and M. Magnani, “Towards effective visual analytics on multiplex and multilayer networks,” *Chaos, Solitons and Fractals*, 2015.
- [9] P. Erdős and A. Rényi, “On random graphs,” *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [10] A.-L. Barabasi and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [11] P. Doreian, V. Batagelj, and A. Ferligoj, *Generalized Blockmodeling*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 2005.
- [12] M. Brusco, P. Doreian, D. Steinley, and C. Satornino, “Multiobjective blockmodeling for social network analysis,” *Psychometrika*, vol. 78, no. 3, pp. 498–525, 2013.
- [13] R. Rossi and N. Ahmed, “Role discovery in networks,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 27, no. 4, pp. 1112–1131, April 2015.