Identifying A Set of Key Actors from Large Social Networks for the Collaborative Performance

R Suhasini[1] , V Swapna[2], G Bhavana[3], Ch.Shanthi[4]

1 Assistant Professor, CMR College of Engineering and Technology

2 Assistant Professor, Keshav Memorial Engineering College

3 Assistant Professor, Keshav Memorial Engineering College

4 Assistant Professor, Keshav Memorial Engineering College

Abstract:

Social network analysis (SNA) becomes one of the certain tools for the accomplishment of the twenty-first century's civilization. Collecting relational data from structured/unstructured documents, network modelling, and obtaining actionable insights need expertizing and awareness in certain fields. Identification of communities, key actors, predicting links/attributes are the key tasks of modern social network analysis..

In the network analysis, the aforementioned assumptions are considered for the development of centrality measures. The centrality measures were proposed to rank the actors. Many works focused towards the improvement of the actor level centrality computation for large social networks. Thus, approximation, parallel, compression based techniques were proposed. Simultaneously, many research considers more than one path, alternate paths for centrality score computations. Proposal for disconnected networks also came out. An extensive study on
related centrality measures can be found in.

Several works were proposed for identifying the influential set of actors. However, we observed that the collaborative performance of the leading K actors identified by actor level centrality measures is not always best. Rather in many cases, an inferior set of actors collectively produces superior performance. Group centrality was proposed to compute collective centrality score of a given group of vertices. However, much work has not carried out for identifying the best group from the huge combinations. For a large network, it is also not practical to investigate all possible combinations and then identify the superior one. Along with this, the group-centrality approach is restricted within the degree, closeness, and betweenness centrality measures only. For the other measures, the group-based approach is still unattended. Techniques such as combinatorial optimization, k-shell decomposition were proposed to identify an influential set of actors. In this paper we are going to find out the set of key actor identification for collaborative performance using machine learning strategy.

Keywords: Social Network, Actor, Centrality measures.

1. Introduction:

Social network analysis (SNA) comprises a collection of techniques concerning the study of social structures using networks and graph theory. The analysis focuses towards exploring patterns of people's interactions. SNA explains the relational strength of networked elements by analysing the networked structures. The relations are analysed with reference to nodes, vertices, or actors (i.e., individual elements such as people, activities, events etc.) and the ties, edges, or links (interactions or relationships) that connect network elements. SNA utilized statistical approaches to model, approximate, tagging, to understand social behaviour, and to predict interactions. It is being used to understand network dynamics. Thus using SNA it is become possible to identify homophily (i.e., similarities of elements and sub-networks), multiplexity (i.e., relationship strength), mutuality/reciprocity, network closure (i.e., to understand completeness and transitivity), propinquity (i.e., geographically closeness), central/influential/prestigious actors, bridges, structural holes (i.e., missing ties between network components), prominent groups, and prediction of relations/attributes for the prosperity and modernization of society. The key tasks of SNA can be classified as follows:

- Social network analysis
    - Network modelling
    - Descriptive analysis
        - Graphical analysis of networks
        - Identification of key actors
        - Community detection
- Predictive analysis
    - Link/attribute prediction

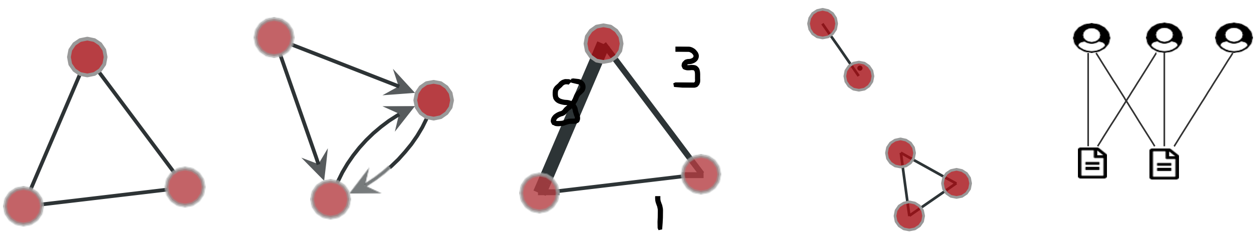
The classified tasks of SNA is described below.

1.1 Network modelling:

The schema of a network model is essentially a graph, in which, the network elements and relationships between elements are depicted as nodes and links respectively. Thus basic relational model consists of two tuples, representing a collection of nodes and collection of links. Generally, a link is a binary relation. Hence they are modelled by two end nodes. In contrast to this, in complex (unusual) form a link may connect many nodes. Nodes and links are attributed with various features. However, based on the relational information and method of analysis, networks are modelled as undirected/directed, unweighted/weighted, connected/disconnected, labelled, multi-partite networks, and modelling of dynamic networks. With the bidirectional links, an undirected modelled network represents symmetric two-way
relational strength among two incident actors. Edges of a directed network indicate the direction of flow between two actors. Thus the relational strength between two actors may be asymmetric.

For example, A votes for B. An unweighted network consist of equal edge weights for all edges, representing equal relational strength. In contrast to this, in weighted network edges carries information of relational strength with respective amount of edge weight. In a connected network there exists at least one direct/indirect connection between any two pairs of actors. However in a disconnected network there exists at least one pair of actors which are unreachable from one another. Nodes and links in a labelled network (a.k.a, social-attributed network) contain various properties, such as the gender, age of actors, occupation etc. The multi-partite network contains more than one category of actors, and a link in the partite network connects two different category nodes. Containing two categories of nodes, bipartite networks are most common in modern social network analysis. Conversion of multi-partite networks such as bipartite networks to a simple network is a convenient approach for the sake of network analysis.

A dynamic network considers time as a network parameter. Temporal and streaming network are the mainly two type of dynamic networks. The temporal network includes the relation duration that exists between two actors. A streaming network is mostly an incremental network, where nodes and links are added over time. Figure 1.2 presents different types of modelled networks.



a) Undirected b)Directed c)Weighted d) Disconnected e) Bipartite

Figure 1.2: Example of modeled network

Real networks from several areas exhibit extraordinary structural regularities, such as power laws, small diameters, consists of communities and sub-communities, and so on. Thus based on the structure of the graph, several graph generation models were proposed such as random graph , Erdős-Rényi model, Barabási-Albert model, Preferential attachment model , Jackson-Rodgers model, Kronecker product graph model (KPGM),block-KPGM to obtain synthetic networks. Figure 1.3 presents sample generated networks.

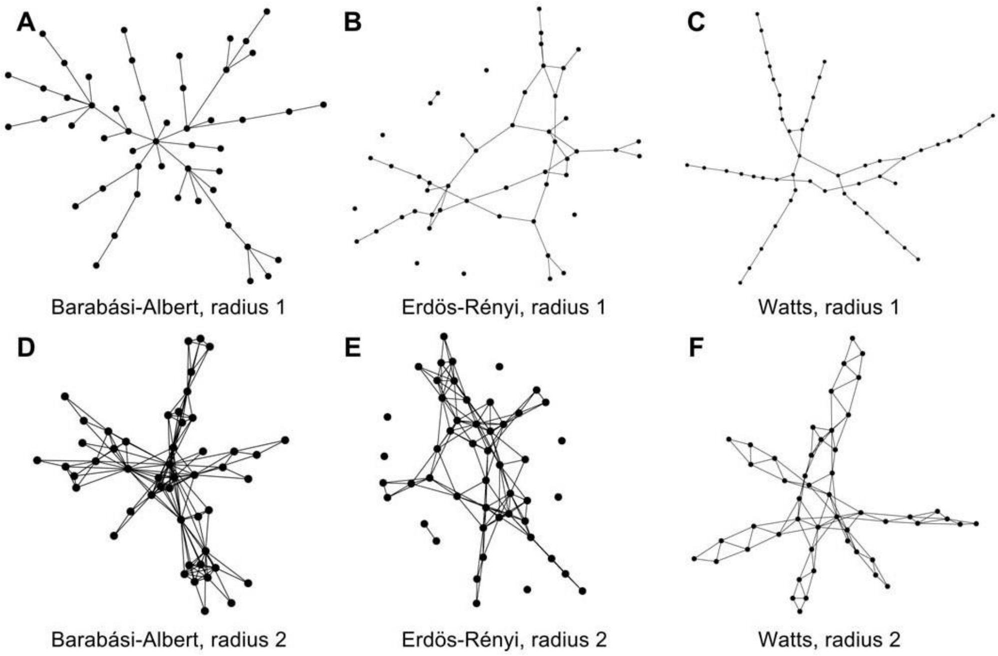


Figure 1.3: Models of network topologies generated using network generators.

1.2 Graphical analysis of networks

In a graphical network analysis, networks are analyzed through network visualization. S*ociograms* are usually used to visualize networks. In sociograms, a point describes an actor and a line (directional or unidirectional) describes a link. Figure 1.4 presents sociogram for a toy network. In this method, it is very easy to identify groups, important actors. However, the graphical analysis works only for small networks.

"The success of a visualization is based on deep knowledge and care about the substance, and the quality, relevance and integrity of the content". A success rate of graphical network analysis depends on:

- Differentiating actors according to actor characteristic;
- Expressing edges by edge weights;
- Minimizing edge crossing;
- Uniform edge length;
- Don't allow nodes to overlap with edges that are not incident on them.

Numerous layout methods such as *circular layout*, *tree layout*, *spectral layout*, *force-directed layout*, *force atlas layout*, *arc diagrams*, *yifan-hu multilevel layout*, *nonverlap layout*, *geo layout*, *layered graph drawing*, *dominance drawing* were proposed for effectively placing of network elements and to improve the visualization. Discuss several network layout techniques.

2. Research Work:

Extending actor level centrality measure to group level, Everett and Borgatti proposed group centrality to measure collaborative centrality score of a group. However, the computation is limited to degree, closeness, and betweenness centrality only. It is also difficult to identify group containing maximum collaborative performance. To find sets of key actors in a social network, Borgatti defined two subproblems for spreadingand obstructing over the network. For efficiently spreading contents over the network, the *key player problem/positive (KPP-POS)* identificatifies key actors by considering key actors as seeds. For the obstructing, *key player problem/negative (KPP-NEG)* progressively removes key actors. The greedy approach starts with an initial set of *k* randomly selected seeds asthe key actors and then keeps swapping one of these actors with another actor outside thisset until no further coverage improvement can be made.

2. This approach does not guarantee a maximum coverage of the network because the coverage of the selected actors could overlap. Using connectivity entropy and centrality entropy Ortiz-Arroyo and Hussain suggesteda measure that utilizes information theory to obtain KPP-POS and KPP-NEG key actorssets. Using Bayes' posterior probability Hussain and Ortiz-Arroyo identified key actors for a given social network. Kitsak et al. ecognized that utmost influential actors are located in the core of networks and to discover such actors authors proposed k-shell decom- position approach by recurrently pruning degree 1, 2, . . . , k nodes. Extending eigenvector centrality, Ilyas and Radha proposed principal component centrality to recognize theset of prestigious actors that are also not near to one another. Chen et al. used a semi- local centrality measure as a trade-off to identify set of influential actors in complex networks and used Susceptible-Infected-Recovered (SIR) model for evaluating influence of actors by using spreading rate and the number of infected actors. Li et al. identify influential spreaders by Leader Rank and Zhang et al. improve the strategy. Liu et al. locate influential actors via dynamics-sensitive centrality. Except group based centrality, KPP-POS, and KPP-NEG, most of the approaches lack to measure collaborative performance. Probst et al. studied state-of-the-art approaches related to key-actor identification. Wu et al.proposed integer programming (IP) formulation of the key player problem (KPP). Authors proposed semi-definite program-integer programming (SDP-IP) and semi-definite program-greedy (SDP-Greedy) algorithm. Gunasekara et al.have applied genetic algorithm (GA) to find the best set of actors for collaborative performance. Authors also identified key actors considering multiple objectives.

3. Methodology:

It is widely believed that user exchanges contents, ideas, and information widely and quickly through the network (a.k.a. word-of-mouth) [225]. The fundamental purpose of the analysis of the diffusion process is to differentiate a set of individuals on the basis of their social ability for information retrieval, manipulation, and propagation. Thus identification of key actors from a given social network is a vital for the prosperity of business/community development.

It is widely used in various disciplines in particular sociology, finance, politics, physics, commerce, healthcare, and in many real-life business applications, such as influence propagation, viral marketing, political campaigning, stop spreading of infectious disease, choosing leaders in management.

In the modern social network analysis, identifying a crucial set of actors is recognized as a very significant problem. The problem of identifying the crucial set of actors is identified as an intractable problem. In last few decades, many actor level centrality measures wereproposed to rank network actors based on the context of study as an entity's position by assigning them with relative ranking.

For instance, the *degree centrality* gives a sense of how the entity has the control of influence in the network with a direct connection. *Closeness centrality* depicts the edge of accessibility within the network. *Harmonic centrality* depicts the edge of accessibility when the network is disconnected. *Betweenness centrality* to check the capability of control of the flow as an intermediary element. *Eigenvector centrality* , *PageRank* to identify prestigious entities. *Katz centrality* all walks among two actors. Thus centrality measures quantify the relative importance of individuals. An comprehensive study on related centrality measures can be found in.

It is observed that, influence-overlapping is a notable issue for a set of key actors identified by an actor level centrality measure, i.e., one or more actor's contribution may overlap. We observe that, the collaborative performance of leading K actors identified by actor level centrality measures is not always best. Rather, an inferior set of actors collaboratively produce superior performance.

3.1 Community detection

Social networks inherently carries "community structure" within it. Communities (also known as clusters or modules) are relatively densely connected subnetworks that also have a weaker set of connections to the rest of the network. It is observed that people divided into groups according to mutual interests, age, gender, occupation, etc. Hence the structure is derived. In a broader sense communities are groups of actors which apparently share common characteristics and/or perform alike roles inside the network.

Since the early 1970's, researchers started exploring communities within networks. Typically, a community detection algorithm tries to split a network into several subnetworks. However, substantially an actor may belong to several communities. In the social network analysis, the circumstance is described as overlapping community. Under particular circumstances, when the focus is on identifying communities concerning specific characteristics, the goal is to recognize communities where the target set of actors are present.

The quality of a community commonly measured by modularity, normalized mutual information (NMI), and multi-criterion scores. Newman proposed modularity as the ratio of edges that fall within the community minus the expected ratio of edges if the edges are randomly distributed. However modularity unable to quantify small communities due to the resolution limit. NMI is usually employed to quantify the community detection accuracy, thus it is possible to know the structure of the underlying communities in advance. Leskovec et al. considers Internal density, Expansion, Conductance, Normalized Cut, Cut Ratio, Average-ODF, Maximum-ODF, Flake-ODF. For the reliable understanding of the communities, Multi-criterion scores are used to assess the communities from multiple viewpoints.

The community detection algorithms are classified into three major groups:

- Traditional algorithms,
- Overlapping community detection algorithms, and
- Local community detection algorithms.

3.1.1 Traditional algorithms of community detection:

Assuming *k* clusters in the network, a partitional clustering algorithm such as k-means, k-clustering sum, minimum k-clustering, k-median, k-center attempts to segregate network elements into *k* components based on distances between elements and the cluster centroid. Without predefined assumptions, a hierarchical algorithm able to describe hierarchical community structure allowing users to determine the communities according to the convenience later. Agglomerative and divisive algorithms are the two types of hierarchical clustering technique. An agglomerative algorithm iteratively combined communities in a bottom-up fashion if they are adequately similarity. Betweenness clustering is a variant of agglomerative clustering. On the other hand, divisive algorithms iteratively split the communities in a top-down fashion by eliminatinglinks between low similarity actors. A hierarchical clustering algorithm produces a dendro- gram, and communities are obtained by cutting the tree. Girvan and Newman proposed a community detection algorithm (G-N algorithm) based on betweenness centrality scores. The algorithm repeatedly discards links having maximum betweenness centrality score and then recompute betweenness centrality scores for every affected link. Authors also introduced modularity Q to quantify community quality. The approach combines two communities with the highest increment in Q and then obtain dendrogram tree. Leicht and Newman enhance modularity to work for directed networks. In planted 1-partition model , a partition is known as one "plants", consist of a certain number of actors. Each actor has a greater interconnection probability p_{in} as compared to a probability p_{out} of being connectedto various communities . GN benchmark is a most popular version of 1-partition model. Adding community size and power-law degree distributions to GN benchmark Lanc Chinetti derive the LFR benchmark. Based on maximization objectives, Kernighan-Lin's greedy optimization algorithm swapes actors between communities. Spectral clustering algorithm employs min cut ratio. With the flow probability Infomap, split networks into modules.

3.1.2 Overlapping community detection algorithms: Besides the crisp community detection, several algorithms were proposed considering the overlapping feature. The Click propagation method (CPM) [is the pioneer for recognizing overlapping communities. The algorithm starts by finding all clicks then merge adjacent k-clicks if they share k 1 actors. The CPMd was proposed by improving CPM algorithm. Based on hierarchical clustering the LINK algorithm computes the similarity of two links by the Jaccard Inbox. Community overlap propagation algorithm (COPRA) improve label propagation algorithm for multilevel propagation algorithm. The algorithm keeps the community identifier and coefficient that enables to recognize overlapping communities.

3.1.3 Local community detection algorithms: As network getting large and complex, identifying communities from the whole network becomes much complicated. Thus several algorithms were suggested to identify local communities. Clauset discussed the difficulties lies in local community detection. Based on the local formation, Raghavan et al. proposed the label propagation algorithm (LPA) to identify communities. Initially, the algorithm assigns a unique label to each actor. Then iteratively update labels to maximally connected neighboring actors until a common label is assigned to all actors belongs to a community. Begins with a set of actors and specific criteria, the local node expansion such as seed set expansion approach progressively expand the set of actors to obtain the community. For the seed expansion, traditional PageRank and Personalized PageRank was adopted by Kloumann et al. and Whang et al. respectively. Actor centrally is used in . Order statistics local optimization method (OSLOM) optimizes the local statistical significance of communities. The algorithm first looks for prominent groups until convergence, then attempt to distinguish the inner construction or potential unions thereof, and finally attemptto recognize the hierarchical formation of the communities. OSLOM is worthy of identifying communities concerning overlapping community, link directions, and community dynamics.

4. Proposed Work:

We realized that the actor level centrality measures are not always handy for the set of actors identification considering collaborative performance. Rather some inferior set of actors collaboratively produces higher performance. Therefore, identifying a set of actors for collective performance is vital. Therefore, following contributions are considered:

- We employ a network translation mechanism for computing the collaborative scoreof a set of actors and utilize biogeography-based optimization for determining the best set of actors.
- We have shown that due to the modification of network structure, network sampling is not appropriate for the set of key actors identification problem.
- It is also experienced that restricting duplication solution generation during evolution is not a fruitful resolution.

- A population initialization, actor selection strategy by assigning selection prob-ability to actors, limiting search space by identifying serviceable actors, cache implementation, and invoking mutation beforehand of migration further improves the efficiency up to tenfold.
- To match the modern big data scenario, a parallel version of the algorithm is developed.

Diffusion process in social networks attracts much attention in both academia and industry. Several research efforts were conducted to uncover the insights from diffusion processes. It is believed that through networks user exchanges contents, ideas, and information widely and quickly (a.k.a. word-of-mouth) [225]. The fundamental purpose of the analysis of the diffusion process is to differentiate a set of individuals by their social ability for information retrieval, manipulation, and propagation. Identifying a set of key actors from a given social network is a vital research problem in many disciplines such as sociology, politics, finance, economics, and healthcare. It also becomes one of the fundamental tools for many real-life business applications such as influence propagation, viral marketing, political campaigning, stop spreading of infectious disease, choosing leaders in management, etc.

Identifying the crucial set of actors is one of the complex problems in modern social net-work analysis. The problem is identified as an intractable problem. In last few decades, many centrality measures were proposed to rank network actors. We observe that *the col- laborative performance of leading K actors identified by actor level centrality measures is not always best*. Rather, an inferior set of actors collaboratively produces superior performance. Along with this, real-life decision making considers multiple objectives at the same time.

For instance, in a view of making a team of 5 brand ambassadors for the growth of business so that the spread of acceptance is well distributed. Thus the problem includes 2 objectives:

1. Identify the prestigious actors so that collaborative prestige should be maximum.
2. Total number of people trusting these prestigious actors should be maximum.

In a multi-objective problem, it is very unlucky to match all the objectives simultaneously. Thus a tradeoff between objectives is used to decide the best solution from a set of non- dominated solutions. In this chapter, we have studied related key actor identification strate- gies.

The rest of the chapter is set out as follows. Sections 6.1 and 6.2 presents techniques to identify a set of key actors for the single and multi-objective, respectively along with their shortcomings.

4.1 Single objective set of key actors identification:

Introduces actor level centrality measure to pick top K actors. Then, groupcentrality for collaborative performance and other influential actor identification methods have been presented. Techniques presented in this section are based on a single objective.

4.1.1 Actor level centrality measures to identify key actors set:

Actor level centrality measures depict the positional importance of an actor to access/directthe network. In this approach, leading actors are recognized as key actors.

A summary of such measures used in complex network analysis has been presented in Table 6.1. Other than degree centrality, all other centrality measures are global in nature, they consider whole network for computing the score and are resource intensive. An extensive study on related centrality measures can befound in.

Table 6.1: Summary of actor level network centrality measures

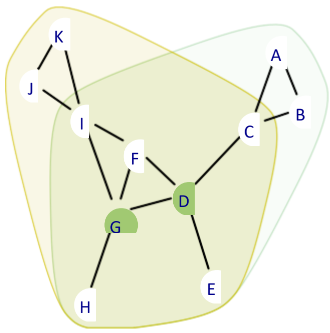
Centrality Measure	Mathematical Definition	Network Type	Description
Degree (C_D) [155]	$C_D(v) = \lvert\Gamma(v)\rvert$	UD, D, UW, W	Directly connected to as many actors as possible.

1

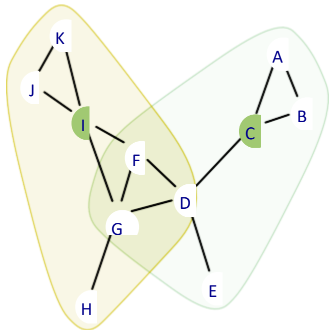
Closeness (C_C) [334]	$C_C(v) = \sum_{u\in V} d(u,v)$	UD, D, UW, W	Should be able to reach to whole network as quick as possible
Harmonic (C_H) [306]	$C_H(v) = \sum_{u \, v\in V} \frac{1}{d(u,v)}$	UD, D, UW, W	Variant of closeness centrality that capable to work on disconnected network.
Betweenness (C_B) [125]	$C_B(v) = \sum_{s\, v/=t\in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$	UD, D, UW, W	Check the capability of control of the flow as an intermediary.
Eigenvector (C_{EV}) [35]	$C(v) = \frac{1}{\lambda} \sum_{u\in V} a_{v,u} x$ W	D, UW, W	Identify prestigious actors. Tend to cluster within a single neighborhood.
Katz (C_K) [189]	$C_K(v) = \sum_{k=1}^{\infty} \sum_{u/=v\in V} \alpha^k (A^k)_{u,v}$	D, UW, W	Instead of only shortest paths, considers influence by taking into account the total number of walks between a pair of actors.
PageRank (C_{PR}) [46]	$C_{PR}(v) = \frac{1-d}{n} + d \sum_{u\in\Gamma_{in}(v)} \frac{C_{PR}(u)}{\lvert\Gamma_{in}(u)\rvert}$	D, UW, W	Another variant of eigenvector centrality adds a scaling factor.

D: Directed network UD: Undirected network W: Weighted network
UW: Unweighted network

Influence-overlapping is a notable issue for a set of key actors identified by an actor level centrality measure, i.e., two or more actor's contribution may overlap. It is also witnessed that the collaborative performance of leading *k* actors identified by this approach are not always best. Figure 6.1 demonstrates the circumstances. In Figure 6.1(a), D and G aretop 2 actors identified by closeness centrality having centrality score of 0.5556 each, and their collaborative score is 0.6923[1]. However, C and I's collaborative closeness centrality score is 0.8181, although their actor level closeness centrality scores are 0.4348 and 0.4545, respectively. Figure 6.1(b) also reveals that for a 2 distance closeness, i.e., maximum walk length of 2 from the given actor – influence-overlapping (intersection of the masked areas) of C and I is significantly less, i.e., 3 vs. 7 or 57% less as compared to D and G in Figure 6.1(a).



(a) D and G are identified by actor level closeness centrality. Yellow and green masks represent masks for vertex D and G respectively. Vertices C, D, E, F, G, H, and I are covered by the influence of both D and G.



(b) C and I are identified by collaborative closeness performance. The yellow and green masks represent masks for vertex I and C respectively. Vertices D, F, and G are covered by the influence of both C and I

Figure 6.1: Problem of collaborative performance for the set of key actors obtained using actorlevel centrality measure as compared to identified using collaborative performance algorithm. Best actors set (green colored) of size 2 concerning network closeness. A background mask indicates influence cover in terms of vertex cover of distance 2 from the corresponding node.

4.1.2 Group centrality for accounting collaborative behavior

Extending actor level centrality measure to the group level, Everett and Borgatti [109] proposed group centrality for computing collaborative centrality score of a given set of actors. Table 6.2 presents mathematical definitions of group centrality measures for a set of actors $\{S \mid S \subset V\}$. However, to the best of our knowledge, the group based definitions are defined for degree, closeness, and betweenness centrality only. Although computation of centrality score for a group of actors is defined for the 3 cases, still determining the best combinationof size K from $\binom{n}{K} = \frac{n!}{K!(n-K)!}$ possible combinations is challenging. For a large network, it is almost impossible to compute all possible combinations and identify the best combination using a traditional von-neumann computer.

Table 6.2: Summary of group centrality measures for a set of actors $\{S \mid S \subset V\}$

Centrality Measure	Mathematical Definition
Degree Centrality	$C_D(S) = \lvert\Gamma(S)\rvert$
Closeness Centrality	$C_C(S) = \sum_{v \in \{V-S\}} d(v, S)$
Betweenness Centrality	$C_B(S) = \sum_{s \neq t \in \{V-S\},\, u \in S} \frac{\sigma_{st}(S)}{\sigma_{st}}$

$d(v, S) = min\,(d(v, u)) \quad \forall u \in S$

$\sigma_{st}(S)$ = Number of shortest paths between s and t that containsany node $u \in S$.

4.1.3 Other methods for identifying set of key actors

Borgatti presents key player problem/positive (KPP-POS) and key player problem/negative (KPP-NEG) to find sets of key actors in a social network. KPP-POS is defined as the identification of key actors for the purpose of optimally diffusing something through the network by using the key actors as seeds. KPP-NEG is defined as the identification of key actors for the purpose of disrupting or fragmenting the network by removing the key nodes. Using connectivity entropy and centrality entropy Ortiz-Arroyo and Hussain have pro- posed an information theory based measure to find KPP-POS and KPP-NEG key node sets. Using Bayes' posterior probability Hussain and Ortiz-Arroyo identified key actors for a given social network. Kitsak et al. observed that most efficient information spreaders are located in the core of a network and by recurrently pruning degree 1, 2, . . . , k nodes they proposed k-shell decomposition method to find such nodes. Extending eigenvector centrality, Ilyas and Radha have proposed principal component centrality to identify the prestigious set of nodes which are also not close to each other. Chen, et al. used a semi-local centrality measure as a tradeoff to identify set of influential nodes in complex networks and used Susceptible-Infected-Recovered (SIR) model to evaluate the performance of nodes by using spreading rate and the number of infected nodes. Li, et al. identify influential spreaders by LeaderRank and Zhang, et al. improve the strategy. Liu et al. locate influential nodes via dynamics-sensitive centrality. Except group based centrality, KPP-POS, and KPP-NEG, most of the approaches lack to measure collaborative performance. Probst,et al. studied state-of-the-art approaches related to key-actor identification.

Approaches other than group-based centrality, KPP-POS, and KPP-NEG are lacking of computing collaborative performance.

4.1.4 Evolutionary approach to identify the best set of actors for collaborative performance

In the preceding sections, we have discussed several ways to compute importance of an actor or a set of actors. However finding the best set of actors for collaborative performance is remain unattended. Gunasekara, et al. have applied genetic algorithm (GA) to find the best set of actors for collaborative performance. Authors considered supernet, a transformed network by merging a set of vertices as the supernode. In Table 6.3, the superiority of collaborative performance of a set of actors identified using EA as compared to top k actor level centrality nodes is presented.

Table 6.3: Collaborative behavior of key actors identified using actor level centrality measure and collaborative approach

Network	Centrality	Top 5 actors picked by actor level centrality	Collaborative Score	Best set of 5 actors for collaborative behavior	Collaborative Score
Dolphin∓	C_D	Grin, SN4, Topless, Scabs, Trigger	0.4035	Beescratch, Grin, Jet, SN96, Trigger	**0.6666**
	C_C	SN100, SN9, SN4, Kringel, Grin	0.5278	Beescratch, Grin, Jet, SN96, Trigger	**0.7403**
	C_B	SN100, Beescratch, SN9, SN4, DN63	0.5793	Jet, Kringel, SN4, SN9, Trigger	**0.7050**
	C_{PR}	Grin, Jet, Trigger, Web, SN4	0.1130	Grin, Jet, Kringel, Trigger, Web	**0.1178**
Prisoners∓	C_D	7, 36, 51, 40, 29	0.4355	7, 15, 36, 46, 55	**0.5968**
	C_C	51, 29, 15, 36, 7	0.5849	7, 13, 36, 46, 55	**0.6667**
	C_B	15, 7, 29, 54, 51	0.7436	7, 29, 40, 46, 51	**0.7586**
	C_{PR}	7, 36, 51, 29, 54	0.0932	7, 15, 36, 46, 55	**0.1121**
US power grid∓	C_D	2553, 4458, 3468, 831, 4345	0.0160	171, 1170, 2606, 3782, 4573	**0.1288**
	C_C	1308, 2594, 2605, 1131, 2606	0.0912	171, 1170, 2606, 3782, 4573	**0.1288**
	C_B	4164, 2543, 1243, 4219, 2528	0.6024	426, 1099, 2543, 2617, 3359	**0.7336**
	C_{PR}	4458, 831, 3468, 2553, 1224	0.0049	205, 831, 2575, 3468, 3783	**0.0072**

∓Refer Table 8.10 for the details of the datasets presented in this table

4.2 Review of multi objective set of key actor identification

Gunasekara, et al.have considered multi-objective set of key actors identification for collaborative performance. The population-based technique first constructs respective super- net and supernode for every solution by combining a set of vertices, then applied NSGA-IIto identify the optimal Pareto front. Authors also have suggested a method to reduce the number of solutions from the obtained Pareto front. The approach essentially utilizes GA tofind the optimal set. However, it is well known that the exploration/exploitation capabilityof GA is trivial. In contrast, BBO is very well known for maintaining balance of explorationand exploitation capability . Thus we have combined the best attributes of BBOand NSGA-II for multi-objective optimization. The non-dominated sorting biogeography- based optimization algorithm has become a valid alternative for multi-objective evolutionary optimization.

5. Summary and conclusion

Many business applications need key actors for the growth of the business. Therefore identification of key actors from social networks is a vital research problem. Thus in this paper, we have reviewed the state-of-the-art approaches for identifying key actors. We observed that the group performance of a set of actors is not always superior. In many cases due to their position in the network, individuals contribution overlaps. Group centrality was proposedto quantify group performance of a given group of actors. The approach is also limited to degree, closeness and betweenness centrality. Evolutionary optimization was used to identify the best actor set for the collaborative performance. Few identification approaches were proposed targeting a specific problem. Similarly, very fewer attempts were made to identify the set of actors considering multiple objectives. Thus designing efficient methods for identifying a set of actors for collaborative performance and multiple objectives is essential. Itis also witnessed that network sampling is not suitable for the case of the set of key actors identification. Therefore, some better approach to handle large network is essential.