Intertwining of Communities, Core and Periphery in Complex Networks

Saisujay Masina, Sathvik Nimmaneni, Navaneeth Gatttagoni, Yayati Gupta, Sanatan Sukhija
Mahindra University, École Centrale School of Engineering
Hyderabad, India

***Abstract*—Community structure and core-periphery structures have been extensively studied in complex networks. However, their coexistence is not well understood. The paper proposes a number of metrics to quantify the influence composition of communities in a network. Shell number of a node calculated by k-shell decomposition algorithm has been used as an indicator of influence. The influential power of communities is also quantified. Most of the communities, specifically influential communities, were observed to possess varying levels of influence. k-shell decomposition have also been applied to the induced subgraph of communities to unravel local cores. It is seen that the local and global cores are neither equal more mutually exclusive sets. The experiments performed to find the correlation are designed using domain-specific properties like shell distribution, community distribution, influence measure, and entropy. One of the major findings is the dense presence of core nodes in a few communities. Further, it is seen that local core is always a subset of the global core for communities that consist of global core nodes.**

***Index Terms*—k-shell decomposition, coreness, community structure, core-periphery structure, entropy**

I. Introduction

Graph theory provides a powerful way to study complex systems of interacting objects. Researchers have studied the properties and organizing principles of such networks from a long time. One of the very popular organising principles in network science is homophily. Illustrated by the quotes, "Like attracts like" and "Birds of a feather flock together", it means that similar people are often seen as friends in a network. As illustrated in Figure 1a, homophily leads to the formation of dense communities [3] in a network which are loosely connected to each other. Such dense communities tend to trap any information diffusing in the network. The information easily diffuses within a community with the help of dense links but can't easily propagate from one community to another [2] because of a very small number of inter-community links.

Another popular organising principles is preferential attachment. Quoted as "rich gets richer" [4], it means that persons having lot of friends tend to attract more friends. Further, such popular persons tend to densely connected with each other. But unlike communities, they are also strongly connected to the remaining network. This leads to the formation of core-periphery structure [5] in a network (refer Figure 1b). It has been shown that the persons located in the core of a network tend to be the super-spreaders [6] of an information.

There have been independent studies depicting communities trapping the information and core superspreading it. However, in the real world, both these forces coexist during information diffusion. To understand the unified impact of these meso-scale structures in information diffusion, it is important to understand their structural co-existence. In this paper, we define a number of metrics to understand the intertwining between both these mesoscale structures in a network.

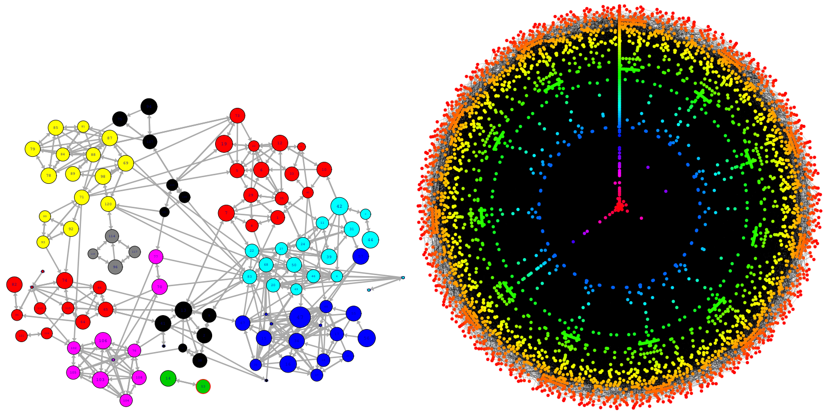
The major contributions of our work are

1) Quantifying the influence of nodes in a network in terms of spreading power and spreading time
2) Proposing a number of metrics to quantify the influential power of communities in a network
3) Understanding the composition of communities in terms of the influence of participating nodes

The rest of the paper is organised as following. Section II highlights the literature. Section III briefly introduces the necessary background work. Section IV to VII propose a number of metrics to quantify the coexistence of two mesoscale structure in networks. Each of these sections consist its own methodology, experiments and results. The paper in concluded in section VIII.

II. Literature Review

The presence of mesoscale structures is observed in most of the real world networks including social networks. Two of these most prominent structures are the community structure and the core periphery structure. It has also been shown that the scale free networks tend to have a core periphery structure [5], [6].These structures affect the pattern of information propagation over these social networks. Communities have been widely observed in real world networks. A number of approaches have been proposed to identify the communities in the network . The seminal work in the direction of a meme virality by Weng et al. [2] emphasizes that the existence of a meme into multiple communities is a signal of a meme getting viral. Another important work by Kitsak et al. [6] proved that there is a dense core at the center of every social network. Once a meme infects a core node, it spreads through the whole of the network. Understanding the spreading of information in different social, biological and economical networks is a problem of high significance. It has been seen that the influence of nodes in a network varies in terms of their spreading power. Hence, identification of such influential nodes remains a major research problem which can help in superspreading/controlling the diffusion of information/diseases in various social, biological and technical



(a) Community Structure (b) Core-periphery structure

Figure 1: [Best viewed in color] Mesoscale structures in a network. a) Friendship network of students in an Indian university b) k-shell decomposition on an induced subgraph on the Facebook network: The red nodes at the boundary are the periphery nodes having least shell number. As we move inwards from the red to the orange, to the yellow nodes; the shell number increases. The red nodes in the centre of the network represent the core of the network.

networks. One of the widely accepted algorithms for determining such influential nodes is k-shell decomposition. It has been used in diverse applications like finding central proteins in PPI (Protein-Protein Interaction) networks, identifying drug targets, understanding mutation rates [7], text visualisation and summarization [8], real-time keyword extraction from conversations [9], corporate networks, banking networks etc. The algorithm is computationally efficient, having a run time complexity of $O(|V|+|E|)$ where $|V|$ is the number of nodes and $|E|$ is the number of edges in the network, respectively. The objective of our work is to study the unified impact of both the above mentioned meso-scale structures on network dynamics. Xiang et al. [10] proposed a single algorithm to predict both the structures. Various theoretical models for the co-existence of both these structures have also been proposed. However, our work aims more at uncovering the relation between these structures in real world networks.

III. Preliminaries

A. Core-periphery structure

The core-periphery structure (refer figure 2a) is a mesoscale network structure, where the network has two types of nodes. The first is a cohesive group of densely connected nodes called core nodes, and the second is nodes that are sparsely connected among themselves are called periphery nodes. There are two types of core-periphery structures. One assumes that a network can have only one core shell, whereas the other model allows multiple core shells. Our paper uses the only first type, which is a network with only a single core shell.

1) K-shell decomposition: The motive of using this algorithm is to find the core shell in the network. In this algorithm, we recursively remove all the nodes with degree ¡= k in the network till no k degree nodes in the network are left and bucket them into the kth shell. We repeat this process for all the values of k (i.e., k=1,2……n) or till the graph becomes empty. At the last iteration of this process, we will get our core shell.

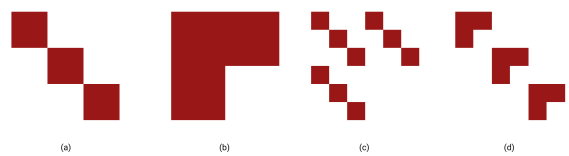


Figure 2: Adjacency matrices of idealized block models. (a) community structure, (b) Core Periphery structure, (c) multiple communities in a core periphery structure (d) multiple core periphery structures in a community.

B. Community structure

It is a type of mesoscale structure network in which a set of nodes are densely connected among themselves. Many algorithms had been proposed to detect such communities (refer figure 2b) automatically.

1) Detecting community structure: There are many algorithms to detect a community structure in a network. Modularity defined as follows can be used to find the most suitable algorithm. Given a network with m edges, the expected number of edges between two nodes i and j(both are of same network) with degrees K_i and K_j respectively is $K_i \times K_j/2m$. $A_{i,j}$ is the actual number of edges between nodes i and j. So, for a network with s communities and a total of m edges, the modularity is given by

$$\sum_{i=1}^{s} \sum_{i \in C_l, j \in C_l} A_{i,j} - \frac{K_i K_j}{2m}$$

A larger value of modularity indicates an excellent community structure. For all the networks used in our work, We have used the multilevel community detection algorithm because of its high modularity. Multilevel community detection algorithm modularity values of the networks can be seen in Table 1.

Network name	Nodes	Edges	Modularity	Number of communities
Karate network	34	78	0.418	4
Facebook network	4039	88234	0.834	16
Politician pages network	5908	41729	0.867	30
Company pages network	14113	52310	0.723	73
Deezer network	28281	9572	0.672	114

Table I: Network properties of Karate, Facebook, Politician Pages, Company Pages and Deezer networks

IV. Information Diffusion Dynamics

A. Cascade Size and Time

Information diffusion in a network is dependent on a plethora of factors. One of the most important and highly researched factor is the set of seed nodes. The choice of the seed set highly impacts the size of the resultant cascade as well as the time taken in the diffusion process. The cascade size and the time resulting from a seed set, $S \in V(G)$, is denoted by $\sigma(S)$ and $\tau(S)$ respectively. Let δ_u represent the shell number of a node $u \in V(G)$. Also, let G_c be the induced subgraph $(V_c, E_c) \subseteq G(V, E)$ for a given shell number c, where,

$$V_c = \{u : u \ \in V(G), \delta_u = c\}$$

and

$$E_c = \{(u, v) : (u, v) \in E(G), u \in V_c, v \in E_c\}$$

. We define the average influential power of a shell G_c as

$$\eta(G_c) = \frac{1}{|V_c|} \sum_{u \in V_c} \sigma(\{u\})$$

. Similarly, we define the average influence time of G_c as

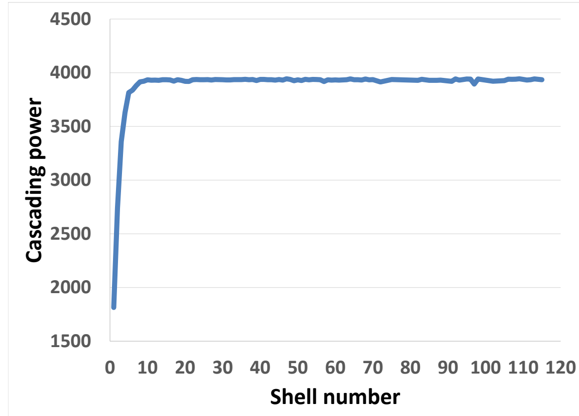
$$\kappa(G_c) = \frac{1}{|V_c|} \sum_{u \in V_c} \tau(\{u\})$$

$\eta(G_c)$ and $\kappa(G_c)$ for various shells in an induced subgraph of Facebook network are shown in Figure 3. As expected, the nodes in innermost shell have the highest cascading power[1]. Such nodes comprise the core of the network.

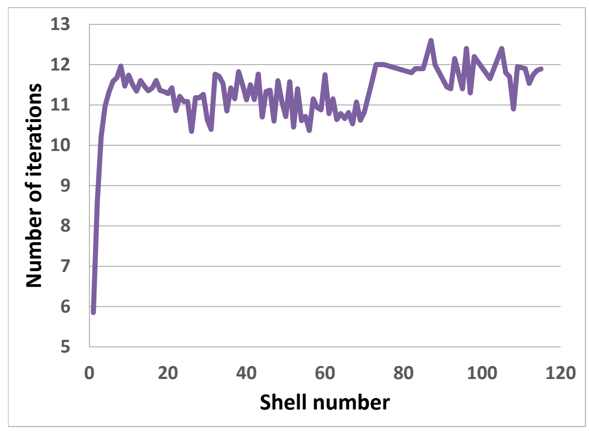
Definition 1. *Core (G^R): It is the induced subgraph on set of nodes comprising the innermost shell of the network.* $G^R(V^R, E^R) \subseteq G(V, E) : R = max(\{\delta_v : v \in V(G)\})$

Figure 3a shows that the cascading power increases exponentially with an increase in shell number. However, it

[1]The terms “influential power”, “cascading power”, “influence”, and “cascade size” have been used interchangeably in the paper.



(a) Cascading Power



(b) Average number of iterations

Figure 3: Average cascading power and Average number of iterations for various shells in the Facebook network.

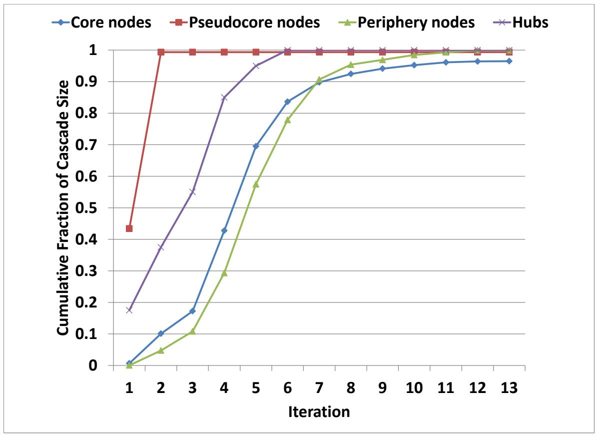
achieves a near maximum value much before reaching R (116 for the Facebook dataset). A similar observation holds true for the cascade time as well (Refer fig). Based on this observation, we define another induced subgraph,

Definition 2. *Pseudo Core ($G^D(V^D, E^D)$): It is the induced subgraph on set of nodes closer to core nodes in terms of influential power.* $G^D = \cup(\{G^x(V^x, E^x) : 1 \leq x \leq R, \eta(G^R) - \eta(G^x) < \Delta\})$, *where* Δ *is a very small value.*

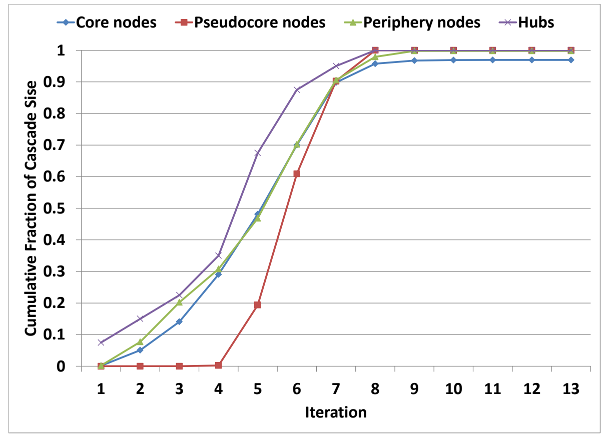
Definition 3. *Periphery (G^P): The remaining network,* $G^P = G - G^R - G^D$, *is termed as periphery. The graph subtraction operation (associative) is defined as follows*

$$G^1(V^1, E^1) - G^2(V^2, E^2) = G^3(V^3, E^3) : V^3 = V^1 - V^2,$$

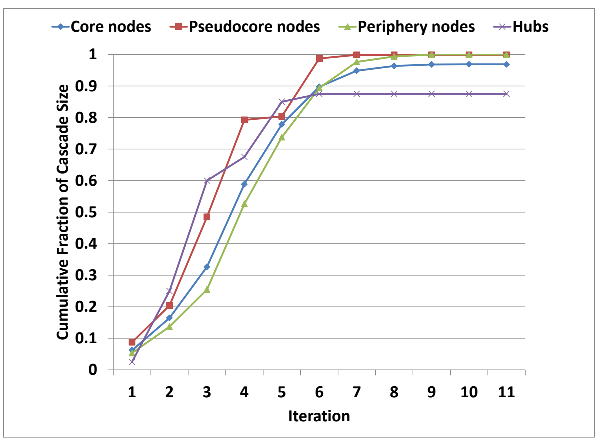
$$E^3 = \{(u, v) : u \in V^3, v \in V^3\}$$



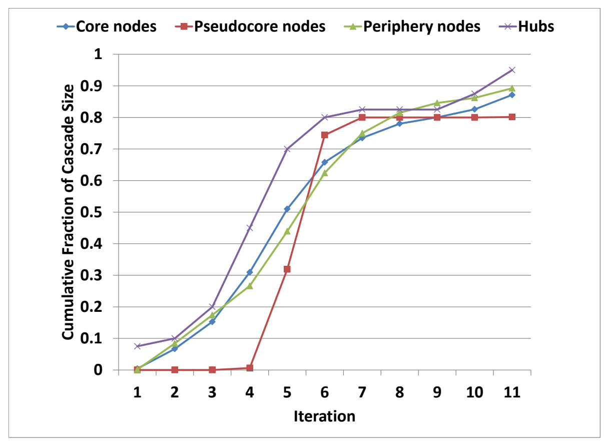
(a) Information cascade starting from core



(b) Information cascade starting from pseudocore



(c) Information cascade starting from hubs



(d) Information cascade starting from periphery

Figure 4: Comparison of cascading power with varying initial seed sets. Top left: $S \subset V^R$, Top Right: $S \subset V^D$, Bottom left: $S \subset H^5$, Bottom Right: $S \subset V^P$. $|S| = 5$ in all the cases.

B. Cascade Growth

We further investigate the change in cascade size over time. For detailed observations, we categorize the nodes of the network, G, in several ways as mentioned below.

1) Core nodes: $V^R(G^R)$
2) Pseudo core nodes: $V^D(G^D)$
3) Periphery nodes: $V^P(G^P)$
4) Hubs (H_k) : H_k represents the set of k highest degree nodes in the network. Considering A to represent the adjacency matrix of the graph G, the degree of a node $deg_u = |\{v : v \in V(G), A_{uv} \in E(G)\}|$.

Next, we consider the seed set S with $|S| = 5$. We simulate 4 experiments with S picking random nodes from the above mentioned sets of nodes. We plot the cumulative number of infected nodes in the sets V^R, V^D, V^P and H_k. Figures 4a and 4c show that the cascade grows exponentially from the beginning of the simulation if initiated from core or hubs. On the other hand; if the seed nodes are pseudocores or periphery (refer figures 4b and 4d), the cascade grows exponentially after a brief initial period. Further in the case of seed nodes being peripheries, the cascade is not complete, i.e., there are nodes in the network which have not adopted the meme.

C. Significance of Communities

The reason why diffusion initiating from periphery nodes leads to a big cascade in our experiments is the random selection of nodes. Given a large number of periphery nodes in the network, it is highly likely for the seed nodes to be picked from structurally diverse regions (different communities) of the network. Weng et al. have shown that cascades initiating from multiple communities are bigger than those initiating from the nodes in the same community. Hence, to understand the information diffusion process in more depth, it is important to collectively study the two meso-scale structures, i.e., core-periphery structure and community structure.

V. Quantifying Influential Communities

To study the intertwining of the community structure and the core periphery struture, we address the following two questions.

- Coreness: What fraction of nodes in a given community are core nodes?
- Core Distribution: What fraction of core nodes are present in a given community?

As discussed before in section III, communities are the dense induced subgraphs in a network. If θ_u represents the community index, m, of a node u, then the corresponding community subgraph $G^m = (V^m, E^m)$, where, $V^m \in \{v : v \in V(G), \theta_v = m, \}$, and $E^m = \{(u, v), u \in V^m, v \in V^m\}$.

A. Coreness of a Community

For each community, $G^m(v^m, E^m)$, we determine

$$f(G^m) = \frac{|\{u : \theta_u = m, u \in R\}|}{|V^m|}$$

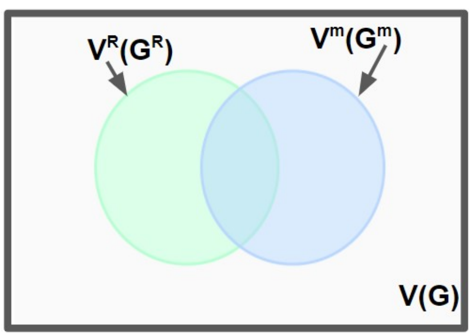
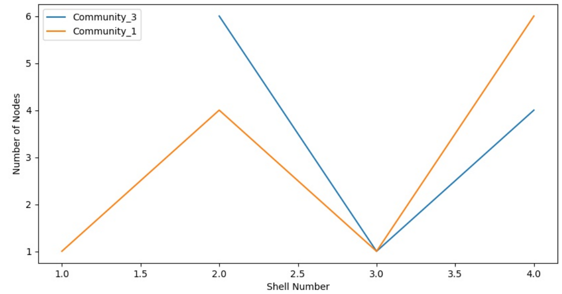


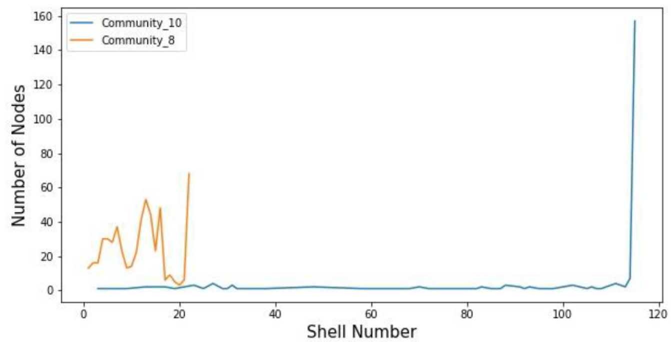
Figure 5: Venn diagram representing the set of nodes V^m in a given community G^m and the set of core nodes V^R. V^G represents the universal set of nodes

Based on the venn diagram representation shown in figure 5, $f(G^m)$ can also be expressed as $\frac{|V^R \cap V^m|}{|V^m|}$.

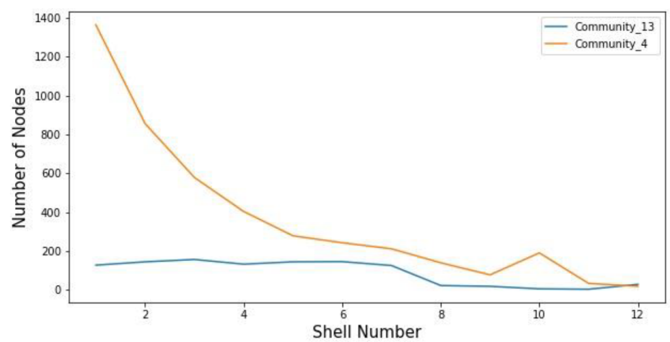
Contrary to our expectations, around 60-70% of the core nodes (in the Facebook dataset) belong to a single community. This shows that the core nodes are densely packed in the same community. However, it is important to verify that this observation is not due to the power law distribution of community sizes in real world networks. We observed that there are a number of large communities comprising mainly of the periphery nodes. On the other hand, there are a few large communities which contain most of the core nodes. The experimental results for the three real world networks are reported in figure 6. X axis represents the shell number and Y axis represent the number of nodes a community has in a given shell.



(a) Karate Network



(b) Facebook Network



(c) Deezer Network

Figure 6: Community size vs. influence. Results of G' and G'' are represented by orange and blue colors respectively.

For each network, we compare two communities,

- the largest community in the network

$$G'(V', E')$$

where

$$|V'| = max(|V^m|,\ \forall m \in \{\theta_u : u \in V(G)\})$$

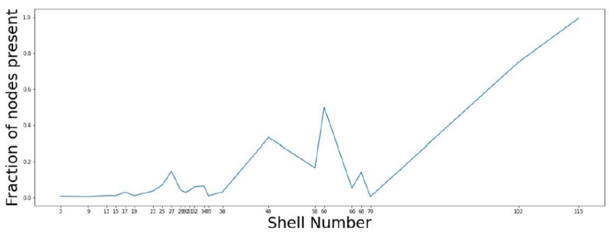
and,

- the community having the largest number of core nodes.

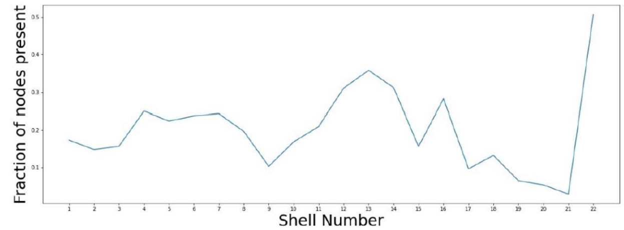
$$G''(V'', E'')$$

where

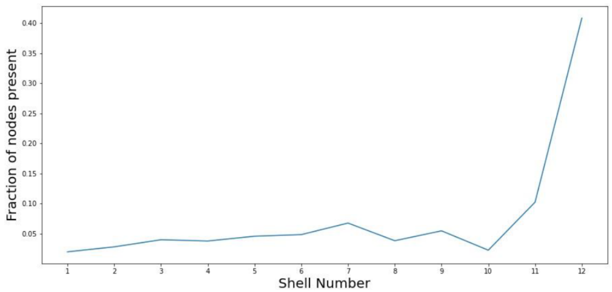
$$|\{v : v \in V'', \delta_v = R\}| = max(|\{w : w \in V, \delta_v = R\}|,$$



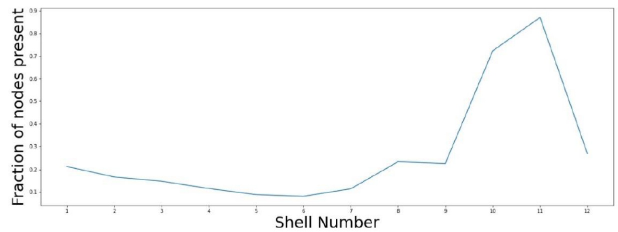
(a) Facebook's community with highest coreness



(b) Largest community of the Facebook network



(c) Deezer's community with highest coreness



(d) Largest community of the Deezer network

Figure 7: Shellwise distribution of nodes in the largest and most influential communities

$$\forall m \in \{\theta_u : u \in V(G)\})$$

.

For comparison purposes; if $G' = G''$, then we modify G'' to be the community having the second largest number of core nodes.

Figure 6 represents three different behaviours shown by different datasets. In the Karate network, the largest community is the most influential one. However, in the Facebook network, it can be clearly seen that the largest community is spread over the initial few shells comprising mostly of the periphery nodes with a few pseudo-core nodes. On the other hand, the most influential community (smaller in size) comprise mostly of the core nodes. $f(G^m) = 70\%$ and $g(G^m) = 99\%$ for the most influential community in the Facebook network. It is seen that influential communities are often large but large communities need not be influential. In the Deezer network, both the largest and the most influential communities are dispersed over multiple shells. However, the largest and most influential communities are different.

B. Core Distribution in Communities

For each community, G^m, we determine the fraction of core nodes present in it as

$$g(G^m) = \frac{|\{u : \theta_u = m, u \in R\}|}{|V^R|}$$

Based on the venn diagram representation shown in figure 5, $g(G^m)$ can also be expressed as $\frac{|V^R \cap V^m|}{|V^R|}$.
Figure 7 represents the number of nodes in various shells for the largest and most influential communities of various datasets.

The largest and the most influential community of the Karate network consist of all peripheral nodes in shell 1 and 60% of nodes from the core. In the Facebook network, the community having the highest coreness also has the highest distribution of core nodes. On the other hand, the largest community contains most of the pseudocore nodes in the network. In the Deezer network, the most influential and the largest communities contains 40% and 30% of the core nodes respectively.
In the above experiments, we mainly focused on the core nodes specifically in the largest and most influential communities. Below, we present more generalised results.

VI. Generalising Community Coreness

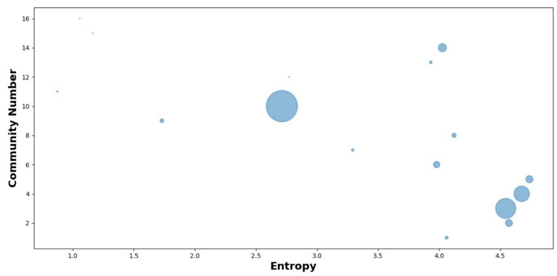
In most of the literature, community and core-periphery structures are studied separately as two different meso-scale structures. Both the communities as well the core are dense subgraphs in a network. However, core is more structurally central in a network. In this section, we consider each G^m where m represents the community index. Then, we independently analyze the influence composition (distribution of various shell numbers) of all communities in the network. For each community G^m, we obtain a vector X^m of shell numbers.

$$\vec{X^m} = (\delta_u : u \in V^m(G^m))$$

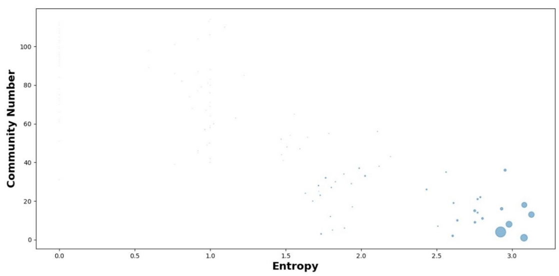
Next, we compute entropy of a community G^m as

$$H(\vec{X^m}) = \sum_{i=1}^{n} P(x^i) log P(x^i)$$

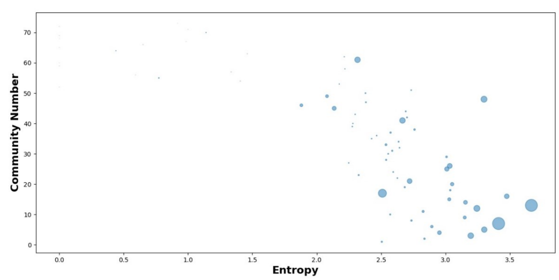
where $n = |V^m|$ and x^i represents i^{th} element of X^m. A high entropy represents a community having nodes of diverse influential powers. On the other hand, a low entropy signifies



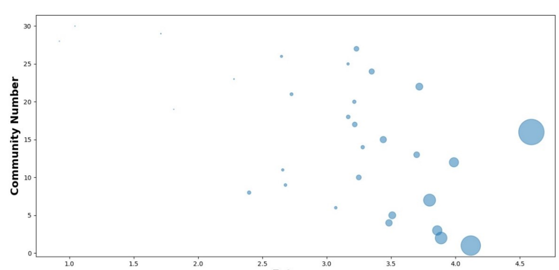
(a) Facebook



(b) Deezer



(c) Company Pages



(d) Politician Pages

Figure 8: Communities' entropy: Each dot represents a community. The size of the dot represents the coreness of the community

equally influential nodes in a community. Figure 8 shows the entropy values of communities in the Facebook network.

It is seen that the most influential communities have high entropy. Facebook community has a highly diversity (entropy) in terms of influence of its participating nodes. However, most of the other influential communities have low entropy values signifying that they are composed mostly of the core nodes. Even though the most influential community has a high entropy, but the entropy of the largest community in the network is greater. Hence, similar to the size, entropy is also not a predictor of the influence of a community.

Since most of the communities comprise a number of nodes other than the core nodes, we generalise the notion of influence as follows. Given a community G^m, its mean influence distribution is defined as $\mu(\vec{X^m})$ as shown below

$$\frac{1}{|\vec{X^m}|} \sum_{x^i \in X^m} x^i$$

Figure 9 shows correlation between the entropy and mean influence of communities in various networks.

In general, a positive correlation can be observed. This means that despite having only a few core nodes, communities with higher entropy are more influential.

VII. Global Influence vs. Community Influence

In the previous experiments, our definition of influence for each node was based on the application of k-shell decomposition to the entire network. However, the algorithm can be applied to the induced subgraph G^m representing the m^{th} community in the network. Hence, now on, we use the term global shell index for δ_u and define another term called local shell index for each node. Local shell index quantifies the influential power of a node within its own community. Local shell number λ_u for a node is computed by applying k-shell decomposition to the community containing the node u. Similar to global core G^R, we define local core for a community G^m as $G^m_Q(V^m_Q, E^m_Q)$ where $Q = max(\lambda_v : v \in G^m)$ and

$$V^m_Q = \{u : u \in G^m, \lambda_u = Q\}$$

$$E^m_Q = \{(u, v) : u \in V^m_Q, v \in V^m_Q\}$$

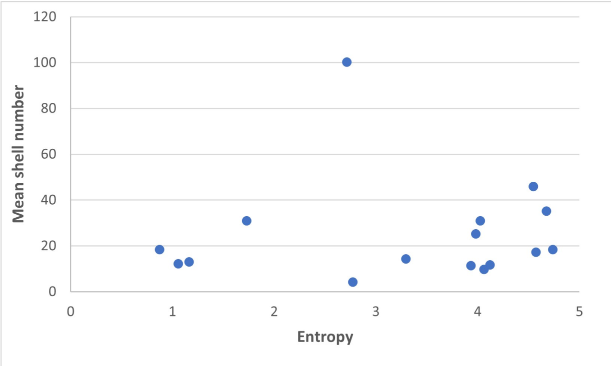
For each community G^m, following metrics can be computed.

1) Jaccard coefficient : $J(V^R, V^m_Q) = \frac{|V^R \cap V^m_Q|}{|V^R \cap V^m_Q|}$
2) Local core fraction: $\frac{|V^m_Q|}{|V^m|}$
3) Global core fraction: $\frac{|V^R \cap V^m|}{|V^m|}$
4) Fraction of the global core in the local core: $\frac{|V^R \cap V^m_Q|}{|V^m_Q|}$
5) Fraction of the local core in the global core : $\frac{|V^R \cap V^m_Q|}{|V^R|}$

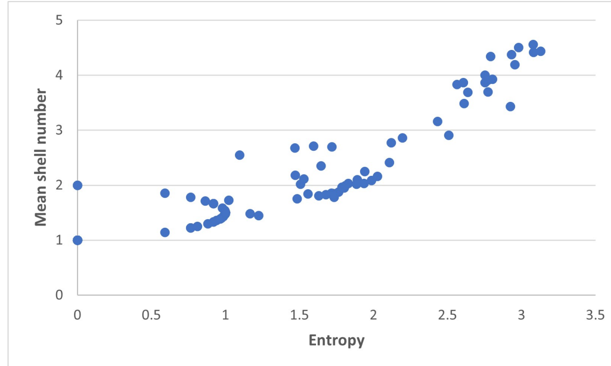
The results of the above metrics for the Karate network are shown in II. From the data, we were able to draw the following conclusions:

VIII. Conclusion and Future Work

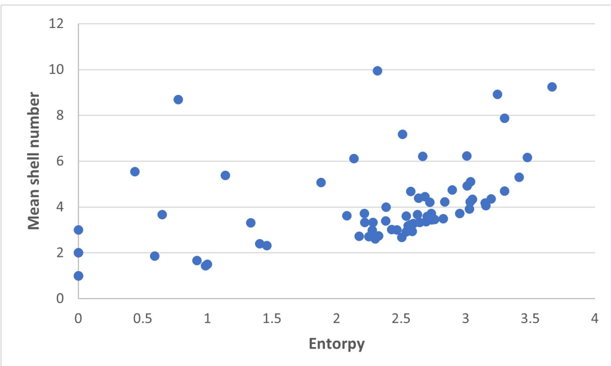
Several metrics have been proposed to quantify the influence of the communities. It has been shown that the community composition and influence are positively correlated. Further, it has been shown that the local and global influence of nodes differ. However, there is a significant overlap between the local and global cores. The presented work motivates a number



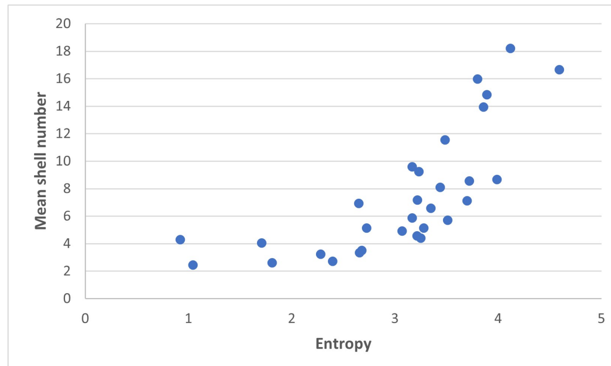
(a) Facebook



(b) Deezer



(c) Company Pages



(d) Politician Pages

Figure 9: The relation between entropy and mean shell distribution shows that as entropy increases, the mean shell of a community also increases.

community number	Jaccard coefficient	Local core fraction	Global core fraction of	Fraction of global core in local core	Fraction of local core in global core
1	0.6	0.5	0.5	0.6	1
2	0	1	0	0	0
3	0.4	0.36	0.36	0.4	1
4	0.83	0	0	0	0

Table II: Metrics for global influence vs. community influence (Karate network)

of metrics to study the intertwining of two most popular mesoscale structures in networks. There is a scope of empirical extension of the study to a large number of networks. As future work, we aim at further refining the proposed metrics to study the differences between global and local core.

References

[1] Zhang, Wenjia, and Jean-Claude Thill. "Mesoscale structures in world city networks." Annals of the American Association of Geographers 109.3 (2019): 887-908.

[2] Weng, Lilian, Filippo Menczer, and Yong-Yeol Ahn. "Virality prediction and community structure in social networks." Scientific reports 3.1 (2013): 1-6.

[3] Fortunato, Santo, and Claudio Castellano. "Community structure in graphs." arXiv preprint arXiv:0712.2716 (2007).

[4] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." science 286.5439 (1999): 509-512.

[5] Borgatti, Stephen P., and Martin G. Everett. "Models of core/periphery structures." Social networks 21.4 (2000): 375-395.

[6] Kitsak, Maksim, et al. "Identification of influential spreaders in complex networks." Nature physics 6.11 (2010): 888-893.

[7] ÇİÇEK, ABDULLAH ERCÜMENT. "k-Shell decomposition reveals structural properties of the gene coexpression network for neurodevelopment." Turkish Journal of Biology 41.2 (2017): 333-341.

[8] Tixier, Antoine J-P., et al. "Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports." Automation in Construction 62 (2016): 45-56.

[9] Meladianos, Polykarpos, et al. "Real-time keyword extraction from conversations." Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017.

[10] Xiang, Bing-Bing, et al. "A unified method of detecting core-periphery structure and community structure in networks." Chaos: An Interdisciplinary Journal of Nonlinear Science 28.1 (2018): 013122.