

RESEARCH

Open Access



Structural hole centrality: evaluating social capital through strategic network formation

Faisal Ghaffar^{1,2*}  and Neil Hurley²

*Correspondence:

ghaffar.faisal@gmail.com

¹ IBM Ireland Limited, Dublin, Ireland

Full list of author information is available at the end of the article

Abstract

Strategic network formation is a branch of network science that takes an economic perspective to the creation of social networks, considering that actors in a network form links in order to maximise some utility that they attain through their connections to other actors in the network. In particular, Jackson's *Connections* model, writes an actor's utility as a sum over all other actors that can be reached along a path in the network of a benefit value that diminishes with the path length. In this paper, we are interested in the "social capital" that an actor retains due to their position in the network. Social capital can be understood as an ability to bond with actors, as well as an ability to form a bridge that connects otherwise disconnected actors. This bridging benefit has previously been modelled in another "structural hole" network formation game, proposed by Kleinberg. In this paper, we develop an approach that generalises the utility of Kleinberg's game and combines it with that of the *Connections* model, to create a utility that models both the bonding and bridging capabilities of an actor with social capital. From this utility and its associated formation game, we derive a new centrality measure, which we dub "structural hole centrality", to identify actors with high social capital. We analyse this measure by applying it to networks of different types, and assessing its correlation to other centrality metrics, using a benchmark dataset of 299 networks, drawn from different domains. Finally, using one social network from the dataset, we illustrate how an actor's "structural hole centrality profile" can be used to identify their bridging and bonding value to the network.

Keywords: Strategic networks, Graph centrality, Social capital, Structural holes

Introduction

With the proliferation of online social networks, the spreading of ideas and information has become easier than ever before. Individuals on these networks connect with each other for various reasons and purposes. In a knowledge-sharing environment, such as professional services or a software development organisation, employees leverage internal social networking platforms to access information in order to solve complex problems. In such competitive environments, individuals who are better at finding information should perform better. It has been shown that when it comes to accessing information and solving problems, people not only rely on their skills and memory, but also heavily on other people [1, 2]. Thus, it can be anticipated that networks that facilitate access to information effectively constitute an important form of social power or value

and contribute to the performance of those engaged in knowledge-intensive work. The underlying structure of a network plays an important role in the spreading and accessing of information [3–5], and studies have shown that certain types of structures are more beneficial than others. For example, Granovetter [1] conceptualised the strength of weak ties (SWT) theory, in which strong ties corresponds to links in a social network over which a high frequency of interactions occur and weak ties correspond to acquaintances. The SWT suggests that weak ties are more likely the source of unique information than strong ties. Similarly, in the enterprise context, studies have found that having more contacts in diverse business units can give access to a wide range of resources that are relevant for the instrumental objectives of career success [6, 7].

Organisations such as enterprises with large work-forces are becoming more interested in understanding the dynamics of networking among employees and in identifying those employees who are key to the flow of information and knowledge through the organisation and those who facilitate innovation and new ideas. Individual workers, on the other hand, are realising the necessity to build and manage their own social contacts in a way that develops their own career prospects. Thus, there is an interest in understanding the nature of social capital, how to detect it and how to curate it. It is in this context that the work of this paper is presented. The paper is focused on social capital, which is understood to depend on an individual's ability to bond with others and to form bridges between diverse groups. In particular, we study social capital through the prism of a new strategic network formation game. Such games have been studied in the state-of-the-art as models of how networks evolve through the actions of nodes choosing their connections in order to optimise some measure of personal utility attained from the network. In particular, we present a game in which the utility corresponds to social capital value. The paper offers two contributions; one in this area of strategic network formation and a second in the area of social capital measurement. Specifically,

1. We propose a new model for strategic network formation that generalises and combines two models from the state-of-the-art and takes into account value accruing to individuals in the network due to both their direct and indirect contacts (i.e. value attained through bonding); and value accruing due to acting as intermediaries between other individuals in the network (i.e. value attained through bridging);
2. from the network formation game, we derive a new measure of social capital—a *structural hole centrality* measure that identifies individuals in a social network whose social connections provide them with bonding and bridging advantages over their peers.

We demonstrate the application of this new measure on a number of networks and carry out a thorough comparison of it, to a number of other well-known centrality measures, using a dataset of 299 networks from different application domains.

The remainder of the paper is organised as follows. In the next section, we review the state-of-the-art on social capital measures and strategic network formation. In “[A bonding and bridging strategic game](#)” section, we develop the new strategic network formation model and in “[Structural hole centrality \(SHCE\)](#)” section, we derive the new “structural hole centrality” measure. Finally in “[Evaluation](#)” section, we present an analysis and

evaluation of the new centrality measure. This section is completed with a case-study of how new measure can be used in practice, in a study of social capital in a network of Norwegian board directors.

Related work

Social capital and its measurement

At its simplest, social capital is the *value derived from social structures*, such as social relationships and social groups, in pursuit of one's goals [8]. Among various definitions of social capital, Putnam's influential work in [8] describes it as: "*Social capital is about the value of social networks, bonding similar people and bridging diverse people with norms of reciprocity*". This definition emphasises the difference between *bonding*, the value obtained through direct friendship links within communities of homogeneous groups of people, and *bridging*, the value obtained by being a social connector between heterogeneous groups of people. Also influential is Burt's *Structural Hole Theory* [2], which focuses on the role an individual plays in a social network and the position that an individual holds relative to others in the network. A structural hole, is a 'gap' in the social network, an absence of connections between different social groups. An individual who can straddle that gap, by forming a bridge between these disparate groups has access to multiple sources of information and the advantage to control the information flow between these groups. It is by bridging such holes, that new innovation and ideas are often generated [9]. Burt refers to this as the social capital of *brokerage*. In fact it is possible (see, e.g. [10]) to distinguish between bridging and brokerage, by recognising bridging as a property of edges in the network, related to the extent to which an edge forms a bridge and brokerage is a node level property, that captures the extent to which a node controls the bridges in the network.

A natural question that arises, then, is how an individual's social capital can be measured? Given the distinction between homogeneous and heterogeneous groups in Putnam's definition, one approach may be to determine the diversity or similarity of social groups, through the attributes of individuals in the network. However, most work has focused on deriving measures of social capital directly and solely from the network structure due to data privacy concerns associated with individual attributes data. This second approach is also the focus of this paper.

A social capital measurement assigns a numerical value to each actor or node in a social network, that represents their social value. When such a value is based solely on network structure, then the social capital value function is a type of network *centrality* measure. Such centrality measures assign value to nodes in a network according to their "importance", where different notions of importance have been adopted. Many centrality measures have been proposed; Oldham [11] studies the similarities and differences between 17 such measures. In the context of social capital, the most notable measures are closeness centrality [12], that measures an individual's average distance to all other nodes in the network and betweenness centrality [13], which measures the extent to which a node lies on shortest paths between other pairs of nodes in the network. Closeness may be considered as a network measurement of bonding, where a node with high closeness centrality is connected directly or along short paths to many other nodes in the network. Betweenness, on the other hand, can be considered as a measure of

brokerage, since a node with high betweenness centrality is a connector on many short paths between other nodes in the network.

Also notable are a number of centrality measures where the centrality value assigned to a node is the corresponding component of the dominant eigenvector of a particular linear map associated with the graph. Such measures arise when the notion of importance is defined recursively, such that a node's "importance" is based on its association with other "important" nodes. Eigenvector centrality [14] is formed from the components of the dominant eigenvector of the adjacency matrix of the graph and correlates well with degree centrality, that values nodes with a large number of connections highly. The Page-rank centrality [15] is formed from the components of the dominant eigenvector of a matrix derived from a random walk over the (directed) edges of the graph. In this paper, we will follow this approach by proposing another linear map, relevant to bonding and bridging capital.

While the above are general network measures that have been applied in many contexts to networks of different types, a number of measures have been proposed specifically for the purpose of measuring social capital in social networks. Burt's *constraint* measure [9], captures the extent to which a node is constrained from being a connector, due to the energy that the individual expends on maintaining a tightly knit neighbourhood of direct contacts. Thus, a node with a high constraint value, is weak in terms of its ability to act as a broker. Everett and Valente [10] discuss a number of other measures of brokerage and propose that brokerage can be calculated as an *induced* centrality measure [16], that is, that a node's centrality of brokerage, can be derived from an edge centrality measure of bridging. In fact, they specifically propose that a node's brokerage centrality be measured as the average edge-betweenness centrality of the edges incident to it.

One approach that an analyst can take is to compute multiple different centrality metrics on a network and to reach a perspective on a node's social capital, through observing its rank when ordered according to these different metrics. Later in this paper, we will propose a new centrality measure for social capital which is parameterised in a way that allows analysts to directly observe how a node's social capital is divided among its bonding and bridging capabilities. This single measure can then be used to characterise different node types in the network, according to the mix of social capital value that they have accrued.

Strategic network formation

The issue of how networks are formed and evolve is another general question in complex network analysis that has received much attention. Processes such as preferential attachment have been argued to lead to the complex network structures that are seen in diverse fields such as sociology, economics, computer science, and biology. Jackson and Wollinsky [17] introduced an economic perspective to network formation, arguing that networks form as a result of actors in the network strategically choosing their connections in order to maximise some personal utility. In particular, they proposed the *Connections model*, in which actors derive value or benefits through connections along direct or indirect paths to other nodes, where this value diminishes with path length. The core idea is that individuals receive benefits from direct and indirect connections

but must bear some cost of maintaining their direct connections. Individuals make a decision about what personal links to maintain, based on a utility which is the difference between the benefit and cost of their connections. The total value of the network is then the sum of the value of each node's utility and an *efficient* network is one in which the total value is maximised. An alternative perspective is to consider firstly the total value of the network, and then to consider an *allocation function* that determines how that total value is distributed to the nodes in the network.

The process of multiple individuals simultaneously seeking to maximise the utility they receive from the network can be modelled in a game theoretic manner where each player's strategy consists of a set of actors to which they want to link. The research question arises as to whether a network formation *game* can reach an equilibrium in which no individual can gain by modifying their links and what sort of networks are equilibrium networks of this game. A number of different network formation games have been formulated (see [18] for a survey). Modelling as a non-cooperative game must address the fact that conceptually two nodes (aka players in the game) must agree to form a link. Much analysis (e.g. [19, 20]) of the Connections model has focused on the weaker concept of *pairwise stability*, where a network is stable when no pair of nodes can increase their utility by agreeing to form a link and no individual can increase their utility by unilaterally breaking a link. An alternative approach, adopted by Kleinberg et al. [21] is to model network formation as a non-cooperative game in which players are allowed to unilaterally form links. The Connections model may be thought of as a bonding game—value is derived through connection—and does not model the value of bridging/brokerage. To capture bridging benefits, Kleinberg et al. [21] proposed a model for network formation which captures the bridging benefits that an *intermediary node* (a node connecting two unconnected nodes) accrues as a connector between these end-points. It is these two models that we generalise and combine to form a bridging and bonding strategic game, in which value is derived from direct and indirect connections, as well as from bridging along paths between other nodes.

Our proposed strategic network formation model is most closely related to that proposed in [22], which also encapsulates benefits from direct and indirect connections that decay with path length, as well as intermediary benefits. In this work, the authors firstly propose a network value function that generalises that of the Connections model. They then propose a class of allocation rules to determine how the total value of the network is distributed to individual nodes. Within this class of allocation rules is the Myerson value [23] that allocates utilities to nodes in such a way that benefits are attributed to nodes for their role as intermediaries. Nevertheless, the parameterised allocation rule that we propose allows for a simple control mechanism for determining the relative weight of bonding and bridging benefits, and, as we discuss later, this model leads to a parameterised measure of social capital.

A bonding and bridging strategic game

Our starting point for developing a strategic game in which players consider their social capital in choosing their network connections, is Jackson's Connections game [17] and the structural hole game proposed in [21]. We first review these two games and then

show how these can be combined into a single model, in which both bonding and bridging benefits are taken into consideration when forming links.

Connections model

The connections model, which we will refer to as CONN was proposed originally in [17] and introduces the following payoff function, representing the utility or value that a player u receives from a network G :

$$\mu_u(G) = \sum_{v \neq u} \delta_{uv}^{d_{uv}} - \sum_v a_{uv} c_{uv}, \tag{1}$$

where d_{uv} is the geodesic (shortest path) distance between u and v , $\delta_{uv} \leq 1$, is the *benefit* obtained from having a direct link to node v , and c_{uv} is the cost of forming a direct link to v . An important characteristic of this model is that only direct links incur a cost to player u , but u can benefit through indirect connections. However, benefit diminishes the further u is from v . When faced with a decision of who to connect with, the player weighs the cost of that direct connection, with the direct benefit, δ_{uv} , together with the indirect benefits obtained along paths through v 's connections. A common setting for analysis of this model is the *symmetric* CONN , in which $\delta_{uv} = \delta$ and $c_{uv} = c$, are constant for all u, v .

The main point to note about the CONN model is that value through the connections is accrued to the sources of those connections. The fact that u has a path to another player v , allows u to reap the benefit of that connection. Intermediary nodes along the path between u and v obtain value through their own connections to v , but they do not obtain any benefit for their role as connectors between u and v . Thus the CONN does not assign value for the role of being a connector in a structural hole and hence cannot be considered to model the utility of bridging social capital.

Kleinberg's structural hole model

A different strategic network formation model that models the payoff of being a connector in a structural hole is proposed in [21]. We will refer to this as the KSH model. The key difference between KSH and CONN is that, in the KSH , the value of indirect paths is assigned to the *connectors* along these paths, rather than the end-points. Thus, if w is a player that forms a length-two path between vertices u and v , i.e. the edges $(u, w) \in G$ and $(w, v) \in G$, then the value δ_{uv} that u would obtain for a connection to v , is allocated to w instead. In an undirected graph, w accrues both v 's value to u , δ_{uv} and u 's value to v , δ_v . More exactly, since there may be many length-two paths between u and v , the value obtained by each intermediary, w , is a monotonically decreasing function of the number of such paths.

The structural hole model is limited only to intermediaries along length-two paths. A constant payoff δ is associated with direct links. An interesting version of the model considers a *Harmonic* intermediary benefit, in which the value that could be obtained by a direct link between u and v , is instead allocated equally among all intermediaries on length-two paths between them. Keeping with the notation of the CONN model, if δ is the value that a direct link between u and v would assign to u , then an intermediary w , obtains the value

$$\beta_w(u, v) = \begin{cases} \frac{\delta}{m_{2uv}} & m_{2uv} > 0 \text{ and } a_{uv} = 0 \\ 0 & \text{otherwise,} \end{cases}$$

where m_{2uv} is the number of length-two paths between u and v , such that u and v are not directly connected. In the undirected case, this becomes

$$\beta_w(u, v) = \frac{2\delta}{m_{2uv}}$$

as the intermediary receives u 's value from v , as well as v 's value from u . Interestingly, for this version of the model, the network as a whole attains the same total value, summed over all intermediaries, from length-two paths, as it would if those end-points were directly connected. The total value of the network, the sum over all node utilities, is then

$$\delta \left(\sum_{uv} a_{uv} + \sum_{uv} (1 - a_{uv}) \mathbb{1}(m_{2uv} > 0) \right) - \sum_{uv} a_{uv} c_{uv},$$

which can be optimised by choosing edges to maximise the total benefit over the number of direct and length-two paths formed by a given cost of direct links.

Value functions and allocation rules

The CONN and KSH games are examples of a *value function/allocation rule* game. A network is formed in which individuals are connected by social links and those interconnections convey on the group as a whole some total productivity or *value*. Given individual utilities, such as those defined in Eq. (1), the value function of the network is given by

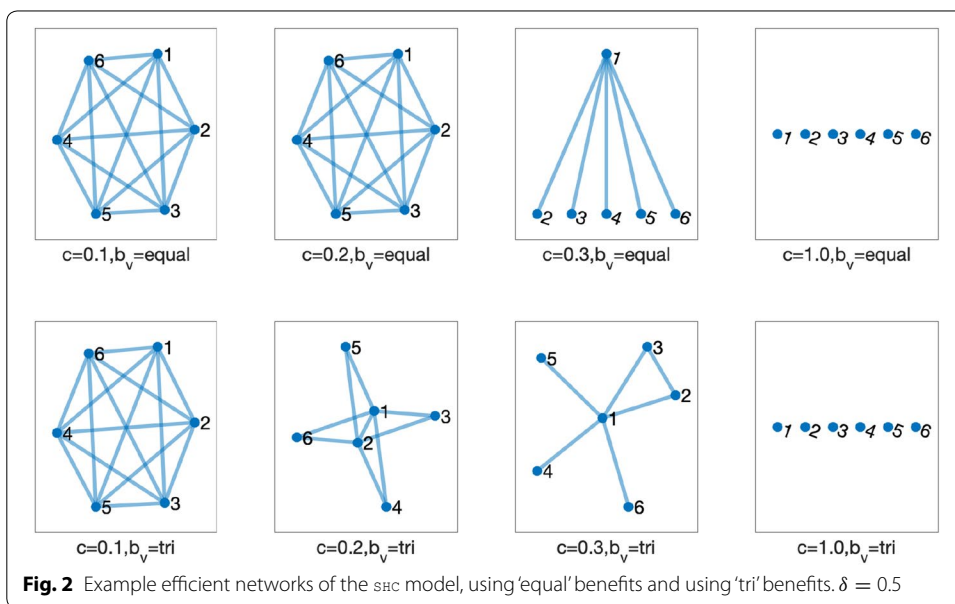
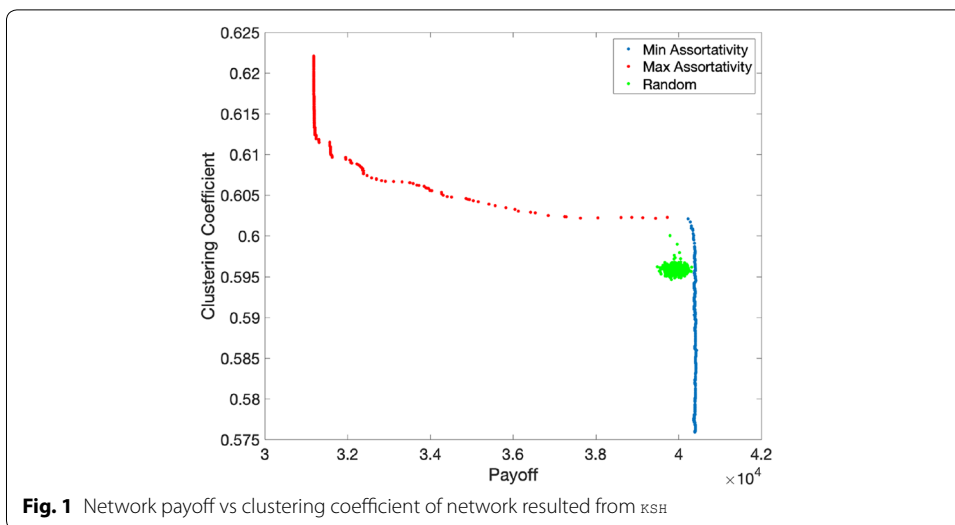
$$\mu(G) = \sum_u \mu_u(G). \tag{2}$$

For a given value function, that assigns a real number to each network over some fixed number of n nodes, it is interesting to consider its *efficient* networks, i.e. those networks that attain the maximum value. For the CONN and the KSH , we have arrived at the value function by summing individual payoffs. Instead, given a value function, it is possible to define an *allocation rule*, that is, a function that distributes the total network value, $\mu(G)$, to the nodes, so that each node obtains a payoff $\mu_u(G)$ such that Eq. (2) holds. It is worth noting, that, for the specialisation of the CONN game in which only length-two paths accrue any benefit, i.e. $\delta_{uv}^{d_{uv}} = 0$ when $d_{uv} > 2$, the CONN and the KSH have the same total value, but it is allocated differently—all benefit goes to the source nodes in the case of CONN , while the indirect benefit goes to the intermediary nodes in the case of KSH .

Limitations of the CONN and KSH

There are a number of limitations to the CONN and KSH models. In particular,

- The KSH model only considers length-two paths for indirect benefits.
- The KSH model allocates the *entire* indirect benefit to intermediary nodes. This eliminates any personal motivation for a player to form indirect links.
- The CONN model allocates *no* benefit to intermediary nodes, ignoring the important role that they play in creating value in the network.



- Neither model takes account of the structural quality of the connecting nodes.

Considering this last point, the efficient networks of the symmetric $CONN$ are studied in [17] and, depending on the relationship between the fixed direct benefit δ and the cost c , consist of either a fully connected network, an empty network or a complete star network, see Fig. 2. In particular, the efficient networks do not contain any triangles, which are known as strong social structures. We argue that the advantage that a node gains from paths in the network, depends on the quality of the end-points of these paths. If the end-points are gateways into strong communities, then there is significant advantage, while if the end-points are themselves dead-ends, or have limited reach into the rest of the network, then they yield relatively less value. We illustrate

this point in Fig. 1. Here, we measure the κ_{SH} value of a network, as the network is modified to increase its clustering. Specifically, starting with a network with a scale-free degree distribution, we carry out pairwise swaps of edges in the network in such a way that the degree distribution remains fixed, while the clustering coefficient of the network varies. The interesting features of this plot are where the payoff remains fixed or nearly fixed, while the clustering coefficient decreases. The reduction in clustering coefficient is indicative of intermediaries in structural holes are connecting between ever weaker community structures. We argue that the payoff of being an intermediary in such a situation should also ideally decrease. We aim to develop a model that accounts for this anomaly and whose efficient networks contain the sort of social structures that we might expect to find in real social networks.

The structural hole connections model $(_{SHC})$

The literature on social capital suggests that an individual's social capital is enhanced by their *bridging* and *bonding* capabilities. The κ_{SH} assigns value to bridging, while the CONN focuses more on bonding, over direct and indirect links. Our goal is to propose a new model, that merges the features of the CONN and the κ_{SH} , to capture both bonding and bridging social capital. We call our model the *structural hole connections model* $(_{SHC})$. In particular,

- We consider the structural value of nodes as the end-points of connections.
- We extend the κ_{SH} to longer paths, maintaining the Harmonic allocation of value to intermediaries on these paths.
- We combine this extended κ_{SH} with the CONN model, so that value is allocated to both source and intermediary nodes along each path.

As will be seen, by maintaining a Harmonic distribution, our extended model retains the same overall value as that of a CONN model and hence our model can be understood as a new allocation function for the value in that model. However, rather than restrict ourselves to the symmetric CONN , instead we consider that the benefit is dependent on the end node, v , so that:

$$\delta_{uv} = \delta^{d_{uv}-1} b_v,$$

where we define b_v as some benefit or value that is obtained through a direct connection to v and the discounting by distance is via a constant δ .¹ The value b_v represents the attraction of forming a connection to player v . For an undirected network, in which edges are bidirectional, the value obtained by u through a connection to v is $\delta^{d_{uv}-1} b_v$, while that obtained by v is $\delta^{d_{vu}-1} b_u$. Hence, each end-point may value the connection differently.

While $b_v > 0$, could capture any type of benefit which may make sense in different contexts, for the purpose of social capital we primarily have in mind, measures of value that capture a node's quality as a connector into a strong community. It is a structural

¹ Note that we are normalising here such that the value of a direct link is 1.

measurement of the neighbourhood v . Such a nodal benefit ensures that the *anonymity* of the network value function is maintained. That is, that the value remains independent of the node labels.² Several such measures are readily available in the complex networks literature. For example,

$$b_v^{\text{tri}} = \sigma_v,$$

where σ_v is the number of triangles that include v as a vertex. Nodes in the network that form many triangles with their neighbours are members of closely knit communities and are hence worth to connect with, either directly or indirectly. Another measure, which considers the density of triangles, rather than a simple count, is the clustering coefficient:

$$b_v^{\text{cc}} = \frac{\sigma_v}{\binom{d_v}{2}},$$

where d_v is the degree of v . Another possible measure is:

$$b_v^{\text{con}} = \sum_w a_{vw} \left(p_{vw} + \sum_u p_{vu} p_{uw} \right)^2,$$

where $p_{vw} = a_{vw}/d_v$. This is Burt's *constraint* measure which captures the extent to which a node is constrained by the community it belongs to. The smaller the constraint, the better a node can act as a structural hole broker. On the other hand, such a broker would like to connect to constrained nodes, as they are members of strong communities.

The KSH assumes that intermediaries connecting nodes u and v , that are not directly connected, receive the value that would otherwise go to the end-points. The value is assigned entirely to the intermediary, while the CONN assumes that nodes obtain value for other nodes to whom they have indirect, as well as direct, connections. In merging these two perspectives, we consider that a source node on a connecting path retains some fraction $\gamma \leq 1$ of the value of the end-point of the path, while the remainder of the value, $(1 - \gamma)$, goes to the intermediaries. Thus, we allocate $100 \times \gamma\%$ of the value, as the CONN does, to the source of the connection and $100 \times (1 - \gamma)\%$ of the value, as the KSH does, to the intermediaries.

Finally, we extend the KSH to longer paths. We retain the Harmonic benefit allocation used in the KSH , so that the full value of an indirect link is retained in the network, but is allocated between intermediary and source nodes. In particular, any intermediary w on a length ℓ geodesic path between u and v , obtains the benefit

$$b_w(\ell, u, v) = \begin{cases} (1 - \gamma) \frac{\delta^{\ell-1} b_v}{(\ell-1) m_{\ell uv}} & 0 < m_{\ell uv}, \ell \leq d_{\max} \\ 0 & \text{otherwise,} \end{cases}$$

² More exactly, given a network $G(V, E)$ defined over nodes $\{v_1, \dots, v_n\}$, and a permutation π of the labels $1, \dots, n$. If $G^\pi(V^\pi, E^\pi)$ is the network such that $(v_{\pi(i)}, v_{\pi(j)}) \in E^\pi \Leftrightarrow (v_i, v_j) \in E$. Then $\mu(G) = \mu(G^\pi)$.

where $m_{\ell uv}$ is the number of geodesic paths of length ℓ between u and v and d_{\max} is some maximum distance beyond which value is lost. (To consider all connecting paths, set d_{\max} to the diameter of the graph; to reduce to the length-two path of the KSH , use $d_{\max} = 2$.) Note that, in this definition, the intermediary benefit is allocated equally to $\ell - 1$ intermediaries along each path over all $m_{\ell uv}$ paths. Also, we have retained the path distance discounting ($\delta^{\ell-1}b_v$) of the CONN model, which was not applied in the original KSH model. To summarise, in the SHC model, a node obtains value from the network

- by direct connections to other nodes;
- by being the source of a length ℓ geodesic path to another node;
- by being an intermediary on a length ℓ geodesic path to another node,

where $1 < \ell \leq d_{\max}$.

In the following, we write the utility of a node w in the graph, by considering these three types of benefit. Firstly, the value obtained by w due to the direct connections can be obtained by summing the nodal benefits b_v over all nodes v that are directly connected to w :

$$\sum_v a_{wv} b_v.$$

Next, the value obtained by w due to being a source of a geodesic path of length ℓ is:

$$\sum_v \left(\sum_{\ell=2}^{d_{\max}} \delta^{\ell-1} \mathbb{1}(m_{\ell wv} > 0) \right) b_v \equiv \sum_v s_{wv} b_v,$$

where s_{wv} is defined by the expression between brackets, which is arrived summing over all possible path lengths ℓ , the discounted benefit obtained by being connected to a node v at the end of such a length ℓ path.

Finally, the value obtained by w due to being an intermediary on a geodesic path of length ℓ is

$$\begin{aligned} & \sum_{\ell=2}^{d_{\max}} \sum_v \sum_u \beta_w(\ell, u, v) \\ &= \sum_{\ell=2}^{d_{\max}} \sum_v \sum_u \sum_{j=2}^{\ell} \delta^{\ell-1} \frac{m_{juw} m_{(\ell-j)wv}}{(\ell-1)m_{\ell uv}} b_v \\ &= \sum_v \left(\sum_{\ell=2}^{d_{\max}} \frac{\delta^{\ell-1}}{\ell-1} \sum_u f_{\ell uwv} \right) b_v \\ &\equiv \sum_v h_{wv} b_v, \end{aligned}$$

where $f_{\ell uwv}$ is the fraction of all length ℓ geodesic paths between u and v that contain w and h_{wv} is defined as the expression above it in brackets, which is arrived at by considering all geodesic paths of length j from a node u to w , followed by all geodesic paths of length $\ell - j$ from w to v .

Now, if we define the matrices $S = \{s_{wv}\}$ and $H = \{h_{wv}\}$ and the cost of connecting to a node v as c_v and the vector of costs as \mathbf{c} . Then the utility vector $\boldsymbol{\mu} = \{\mu_w\}$ for the SHC can be written as:

$$\boldsymbol{\mu} = \left(A + \gamma S(\delta, d_{\max}) + (1 - \gamma)H(\delta, d_{\max}) \right) \mathbf{b} - A\mathbf{c}, \tag{3}$$

where A is adjacency matrix. Note that we have represented the dependence of S and H on the parameters δ and d_{\max} . Observe that $\sum_w s_{wv} = \sum_w h_{wv}$, confirming that the application of either matrix yields the same total value to the network, but the total value is distributed differently by each matrix. In particular, it follows that $\|S\|_1 = \|H\|_1$.

We note that scaling the value vector \mathbf{b} by any multiple does not change the relative value of one node over another. However, if b^{tri} is used, then the benefit values are integers in the set $\{0, \dots, \binom{n-1}{2}\}$, but if b^{cc} is used, then the values are rational numbers in the interval $[0, 1]$. It makes sense therefore to scale \mathbf{b} to a single range of values. In the remainder of this paper, \mathbf{b} is normalised so that, whenever $\mathbf{b} \neq 0$, it contains the same total benefit as the symmetric CONN model, where $\forall v, b_v = 1$. That is:

$$b_v \leftarrow \begin{cases} n \frac{b_v}{\sum_u b_u} & \text{provided } \max_u b_u > 0 \\ 0 & \text{otherwise} \end{cases}$$

We can then scale the cost, to compare different cost/benefit tradeoffs without having to account for any scaling issues due to the choice of the nodal benefit function.

Discussion

While it is beyond the scope of this paper to explore in detail the efficient networks of the SHC, in Fig. 2, we show some example efficient networks for $n = 6$, for the case of a constant benefit $b_v = b_v^{\text{equal}} = 1$, corresponding to a symmetric CONN model, and when $b_v = b^{\text{tri}}$. It may be observed that, in the second case, efficient networks containing triangles are found, as only nodes connected to triangles have a non-zero nodal benefit. This shows that the SHC, with $b_v = b_v^{\text{tri}}$ yields a richer set of efficient networks than the symmetric conn and that they contain structures that are commonly observed in real social networks.

Structural hole centrality (SHCE)

Our primary interest in this paper is to use the SHC game as a means of defining a structural hole centrality measure that can identify nodes in a social network with high social capital.

In the derivation of the SHC, the γ parameter controls the allocation of value to nodes in the network. Different values of γ may be considered as different allocation functions, that distribute the total network value, which is determined by δ, d_{\max}, b_v and the cost c . This total network value is obtained as a sum over all the paths in the network of the path-length discounted benefits obtained from end-points of those paths. The question of a fair allocation of such network value has been addressed in works such as [24]. One approach is to identify desirable properties of the allocation function and determine an allocation that satisfies those properties. Two desirable properties of a fair allocation are that it be *component*

Table 1 shce centrality parameters

Parameter	Description
b_v	Benefit associated with connecting to a node v , where that benefit captures the structural quality of v in the network
δ	Indirect path benefit discount, such that a path of length $\ell \geq 2$ to a node v , accrues a benefit of $\delta^\ell b_v$
d_{\max}	The maximum path length, such that there is no benefit to being connected along a path of greater length
γ	Proportion of the indirect path benefit that is associated to the <i>source</i> of a path
η	The scaled cost of a connection to a node

additive, that is, the value generated by any connected component in a network should be allocated among the nodes in that component; and that it satisfy *equal bargaining power*, that is, that if two nodes, u and v are connected, then the change in the value allocated to node u when the edge (u, v) is removed, should equal the change in the value allocated to node v . Equal bargaining power says that the pair of nodes each benefit or suffer equally from the addition of a link between them. These two properties hold if and only if nodes are allocated their so-called “Myerson value”, defined as:

$$\mu_u(G) = \sum_{W \subset V \setminus \{u\}} \left(\frac{|W|!(n - |W| - 1)!}{n!} \right) (\mu(G|_{W \cup \{u\}}) - \mu(G|_W)),$$

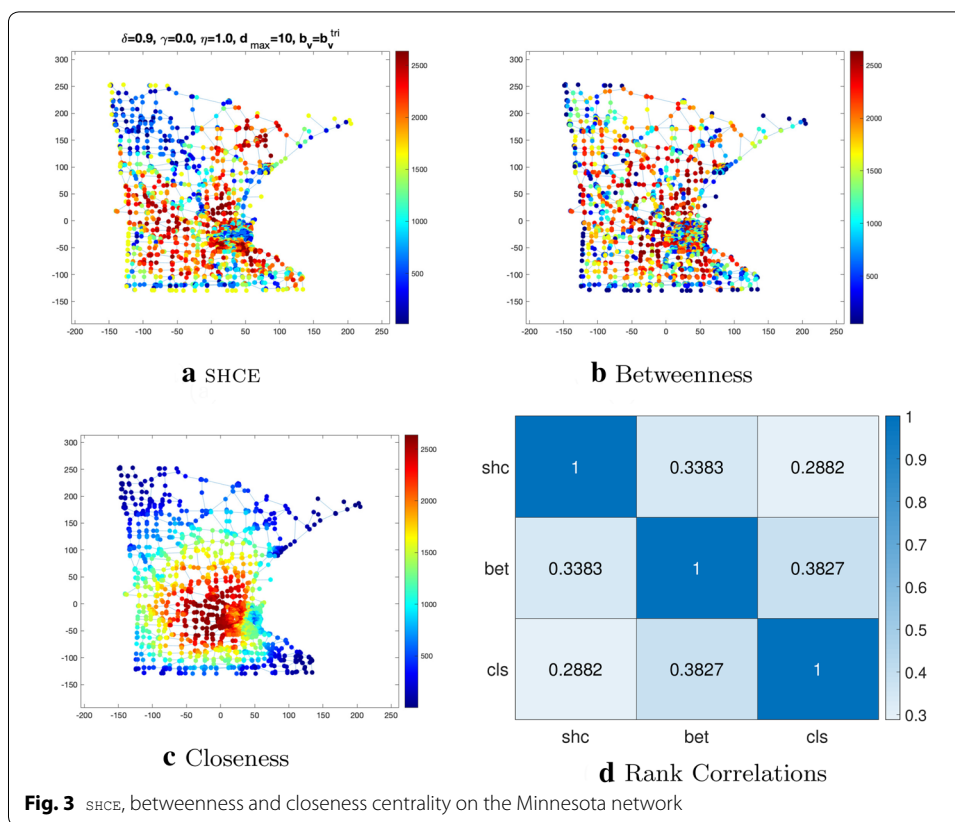
where the sum is over all subsets W of the nodes in the network not containing u and $G|_W$ means the network restricted only to those nodes in W . The Myerson allocation will often allocate high value to intermediate nodes, as they are crucial for the creation of value on paths that traverse them. However, it is not tractable to compute the Myerson value on real-world networks, since the sum is over 2^{n-1} possible subsets.

Instead, the *shc* game allows for the exploration of a range of different allocations, by modifying the value of γ and, when $0 < \gamma < 1$, all nodes along a path get allocated some proportion of the value that is generated by that path. In fact it is generally the case, that the Myerson value correlates strongly (in rank order) to the node utilities of the *shc* for some value of γ , typically when $\gamma \approx 0$. On the other hand, modifying γ allows an analyst to explore how different nodes benefit from different allocation strategies and this can give some insight into their position of influence in the network: when $\gamma \approx 1$, nodes that are connected along short paths to many other nodes can expect to benefit from a high payoff, while when $\gamma \approx 0$, nodes that are intermediaries on many short paths can expect a high payoff. Hence we define the *structural hole centrality* measure, *shce*, as the payoff of the *shc* game. To parameterise the cost, we stick with a fixed cost c for every link, and note that the total value in the network is zero when

$$c \mathbf{1}^T A \mathbf{1} = \mathbf{1}^T (A + S) \mathbf{b}.$$

Hence, we define *shce* as

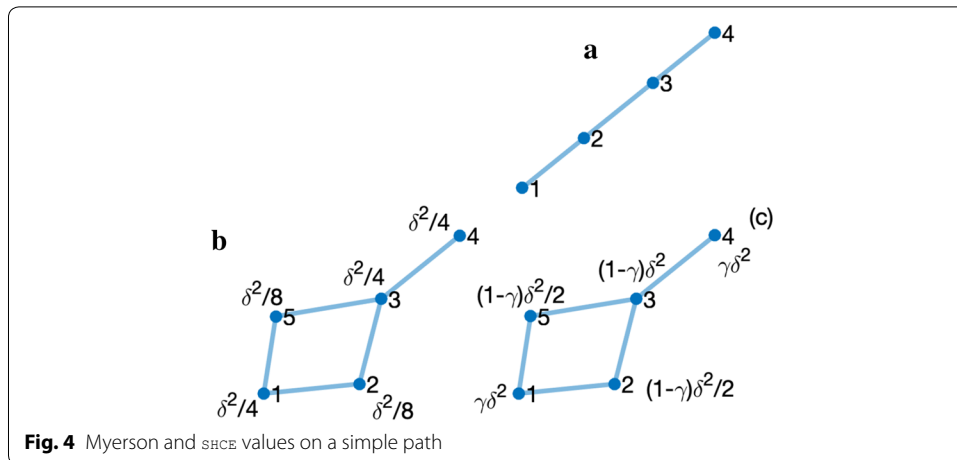
$$\text{SHCE}(\delta, d_{\max}, \mathbf{b}, \gamma, \eta) = \left(A + \gamma S + (1 - \gamma)H \right) \mathbf{b} - \eta \left(\frac{\mathbf{1}^T (A + S) \mathbf{b}}{\mathbf{1}^T A \mathbf{1}} \right) A \mathbf{1},$$



where $\eta \in [0, 1]$, allows the exploration of costs ranging from a zero-cost model to a cost that reduces the value of the network to zero. The parameters of the SHCE are summarised in Table 1.

Relationship to other centrality measures

In “Evaluation” section, we carry out a detailed comparison of the SHCE with a set of other commonly used centrality measures in network analysis. From the above presentation, it is clear that the SHCE is similar to *closeness centrality* (a measure of the average closeness of a node to other nodes in a network) when $\gamma = 1$ and is similar to *edge-betweenness centrality* (a measure of the extent to which a node is found on shortest paths in the network), when $\gamma = 0$. Nevertheless, the SHCE is not identical to either measure. In fact the d_{max} and δ parameters allow for a restriction in the horizon over which a node’s distance to other nodes influences its SHCE value, while closeness and edge-betweenness consider the relationship to *all* nodes in the network. The γ parameter, then allows for a mixture of the betweenness and closeness perspectives. The difference in the measures is illustrated for the Minnesota road network, shown in Fig. 3, which has a diameter of 98. The settings of the SHCE focus value strongly on intermediate nodes, by taking $\gamma = 0$, along with a maximum cost of $\eta = 1.0$ for edges. The plot shows the tied rank of the measures, where nodes with largest centrality value have rank n and nodes with smallest have rank 1. The Spearman rank correlation of SHCE with closeness and betweenness is not particularly strong for these settings. The SHCE also has similarities to the Katz centrality measure, which computes a node’s centrality in



relation to its discounted distance to other nodes in the network. However, the Katz allocates its value solely to the source nodes on such paths and so cannot be used as a measure of bridging capital. We will show in our case-study in “Evaluation” section that computing a profile of SHCE centrality scores as γ is varied allows for some insight into the mix of values that actors get from the position in a social network and provides a single framework with which social capital can be assessed.

Comparison of SHCE with Myerson value

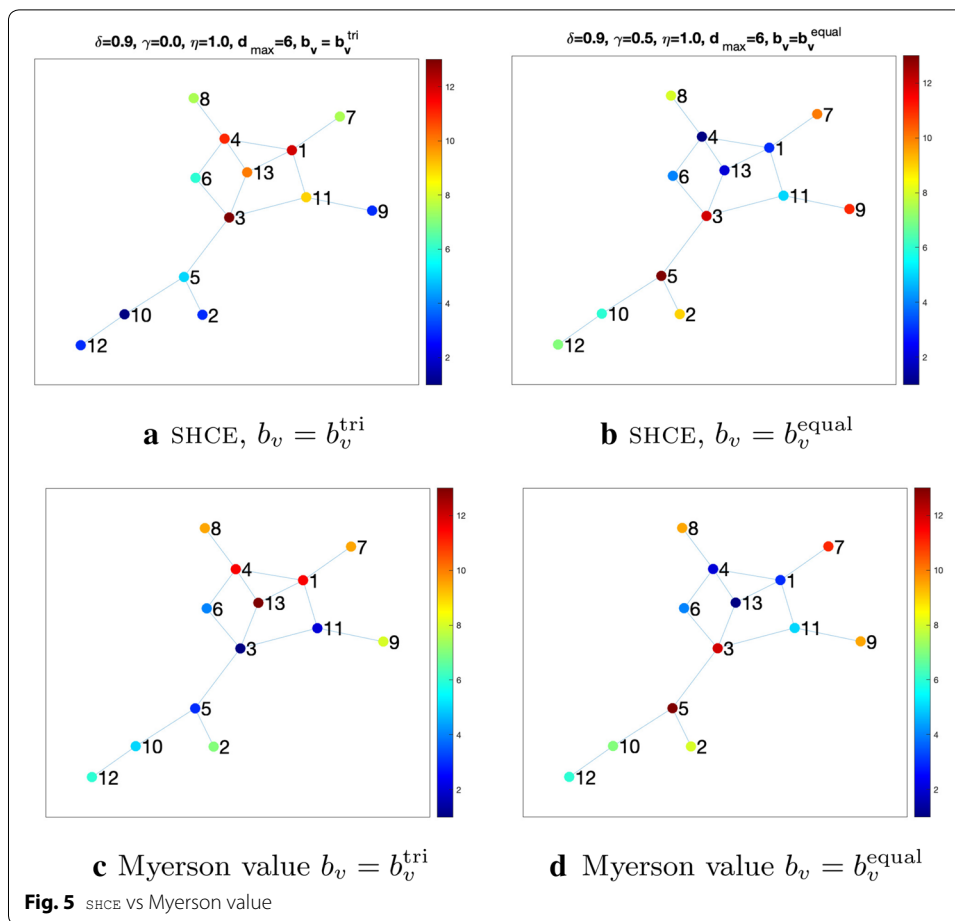
It is instructive to compare the SHC allocation of value to that of the Myerson value in a simple network, with a constant benefit function. In Fig. 4a, we show a network consisting of a single 4-node undirected path. By counting all shortest paths in this network, we can find the total network value as

$$\mu(G) = 2(3(1 - \eta) + 2\delta + \delta^2),$$

which may be observed by counting 3 direct connections (with cost η), 2 paths of length 2 and a single path of length 3, which occur with multiplicity 2, considering that edges are bidirectional. The Myerson value allocates the value of each path evenly among all the nodes along the path, since each node is equally responsible for bringing that value to the network. Hence, each node is allocated the value $2\delta^2/4$ for its contribution to the length 3 path, the end-points receive a value of $\delta/3$ for each of the two length-two paths that start or terminate at them and so on. We can arrive at the Myerson allocation as

$$\mu^{\text{myerson}}(G) = \begin{bmatrix} 1 - \eta + 2\delta/3 + \delta^2/2 \\ 2(1 - \eta) + 4\delta/3 + \delta^2/2 \\ 2(1 - \eta) + 4\delta/3 + \delta^2/2 \\ 1 - \eta + 2\delta/3 + \delta^2/2 \end{bmatrix}.$$

On the other hand, the SHCE allocation depends on the value of γ and is given by



$$\mu^{\text{SHCE}}(G) = \begin{bmatrix} 1 - \eta + \gamma\delta + \gamma\delta^2 \\ 2(1 - \eta) + (2 - \gamma)\delta + (1 - \gamma)\delta^2 \\ 2(1 - \eta) + (2 - \gamma)\delta + (1 - \gamma)\delta^2 \\ 1 - \eta + \gamma\delta + \gamma\delta^2 \end{bmatrix}.$$

If a fifth node is added in order to produce a second length-three path connecting the end-points, as shown in Fig. 4b, c, then the Myerson will distribute the $2\delta^2$ value of that path as shown in Fig. 4b, while the SHCE will do so as shown in Fig. 4c. Both methods give higher weight to node 3 than nodes 2 and 5, since the value remains in the network if either one of these is removed. However, the value that SHCE gives to the end-points of paths depends on the γ parameter.

For further comparison, we examine the relationship between the SHCE and the Myerson value on a random network of $n = 13$ nodes, using both the triangle nodal benefit (b_v^{tri}) and constant nodal benefit (b_v^{equal}) functions. Again, in Fig. 5, the colour indicates the rank of the node. In the case of the triangle benefit, value is concentrated on the nodes that form the single triangle in the network (nodes, 1, 4 and 13), for both measures. The Myerson gives higher values to peripheral nodes 7 and 8, since these nodes add to the value of the network by linking to nodes with non-zero benefit. With $\gamma = 0.0$,

Table 2 Networks from ICON

Domain	Number	Percent
Biological	117	39
Economic	12	4
Informational	17	6
Social	84	28
Technological	53	18
Transportation	16	5

the $SHCE$ focuses more value on intermediary nodes such as 3 and 11 that straddle paths to the non-zero benefit nodes. The overall rank correlation of the $SHCE$ and Myerson is 0.34 in this case. When all nodes have the same benefit, high Myerson values attach to nodes 3 and 5 that add value to the network by forming the path that connects the nodes in the lower left corner to the rest. But, again Myerson credits the peripheral nodes 7, 8 and 9 because they too add to the overall value in the network. The highest correlation (0.99) between these Myerson scores and the $SHCE$ score occurs when we choose a value of $\gamma = 0.5$, that allocates the value equally between source and intermediary nodes on connecting paths. From these examples, it is clear that there is no best value of the $SHCE$ parameters, in the fairness sense from which the Myerson is derived. But it is also generally the case that some settings of the $SHCE$ parameters can achieve centrality scores that correlate strongly with the Myerson. While adjusting γ cannot lead to a fair allocation in the sense of the Myerson value, it can allow insight to be derived into which nodes benefit, when the allocation of value favours bridging capital over bonding or vice versa. The $SHCE$ relies on the analyst to determine an insightful allocation of the value in the network by adjusting its parameters, while the Myerson provides a single best allocation in some well-defined sense. We note however that work such as [24] argues that the fairness criteria of the Myerson may not be appropriate, depending on the context in which the strategic game is analysed.

Evaluation

$SHCE$ correlation with other centralities

The work of [11] is the most comprehensive recent study of network centrality measures. This work examines the correlations between 17 centrality measures across a large range of different graphs, drawn from different application domains. According to this work, the general observed trend for most networks is a “high and positive correlation” between centrality metrics, although there is also “considerable heterogeneity”. To obtain a good understanding of where the $SHCE$ fits in relation to other metrics, it is worthwhile applying this same analysis to the $SHCE$.

Following the work of [11], we evaluate the proposed $SHCE$ centrality measure using a subset of the CommunityFitNet corpus of networks [25] which, in total, contains 572 real-world networks drawn from the Index of Complex Networks (ICON) [26]. The CommunityFitNet corpus includes a variety of network sizes and structures. Our analysis assumes unweighted, simple, undirected networks. We only consider networks with a single connected component and also reject any other networks for which any of the

Table 3 Node centrality measures

Measure	Symbol	Formula
Shortest-path betweenness centrality	bc	$c_w = \sum_{u \neq v \neq w} \sum_{\ell} \mathbb{1}^{\mathbb{K}}(m_{\ell uv} > 0) f_{\ell uvw}$
Shortest-path closeness centrality	cc	$c_w = \frac{n-1}{\sum_v \ell_{wv}}$
Eigenvector centrality	ec	$c_w = \frac{1}{\lambda_1} \sum_v a_{wv} e_v$
Katz centrality	kc	$c_w = \sum_v ((1 - \delta A^T)^{-1} - 1)_{ww}$
Degree centrality	dc	$c_w = \sum_v a_{wv}$
Harmonic centrality	hc	$c_w = \frac{1}{n-1} \sum_v \frac{1}{\ell_{wv}}$
Constraint centrality	conc	$c_w = \sum_v a_{wv} (p_{wv} + \sum_u p_{wu} p_{uv})^2$

Note that ℓ_{wv} is shortest path distance between w and v , δ is the benefit of a direct link, such that a connection along a length ℓ path gets benefit δ^ℓ , e_v is the eigenvector of A corresponding to the dominant eigenvalue λ_1 ; $m_{\ell uv}$, p_{wv} and $f_{\ell uvw}$ are as defined in “A bonding and bridging strategic game” section

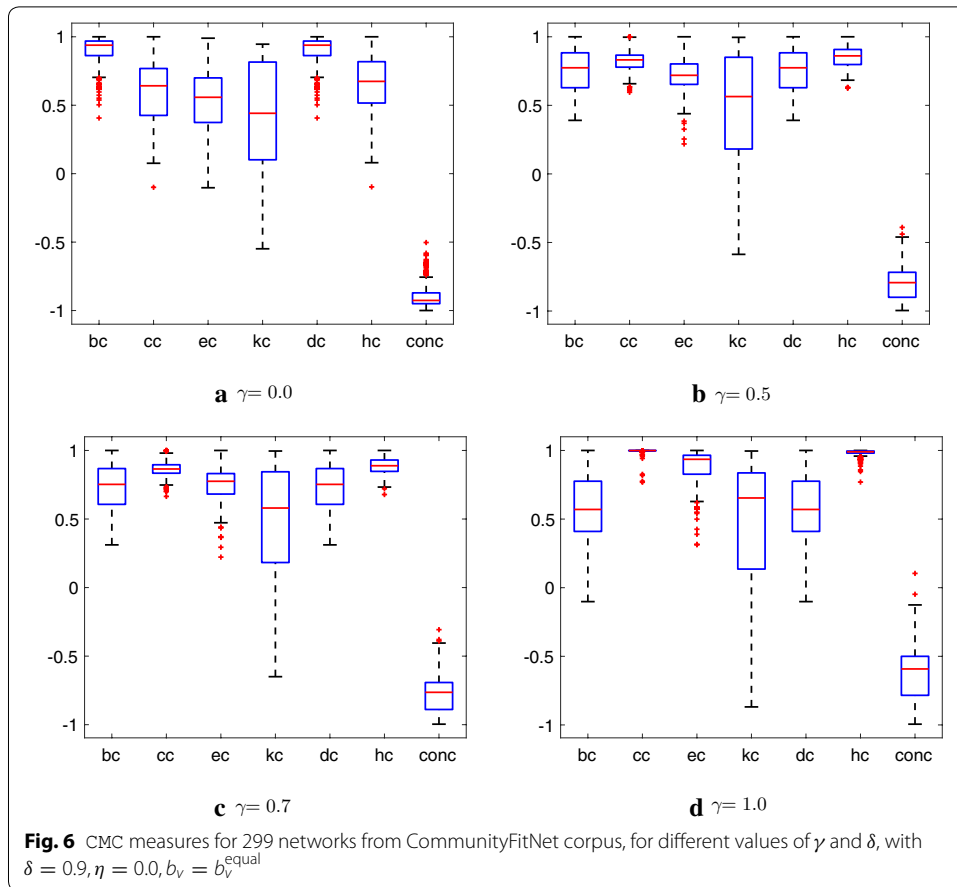
analysed centrality measures fails to compute.³ There remains 299 networks, on which our analysis is performed, which come from 6 different domains (see Table 2), with a range of nodes from 8 to 3155 (average 464) and a mean sparsity of 4.72%.

Following [11], we use Spearman’s ρ as the centrality measure correlation (CMC) between the $SHCE$ measure and various other node centrality measures. This statistic is chosen in [11] on the basis that relationships between measures can be nonlinear, though they are generally always monotonic. The centrality measures that we compare against are listed and defined in Table 3. From their definitions, the connections to the $SHCE$ are apparent. In particular, $SHCE$ relies on values measured along shortest paths, similarly to the cc , hc , bc and kc . Like the cc and hc , path contributions are inversely proportional to their lengths. Like the kc , the contribution of a path decays according to a benefit factor $\delta < 1$, such that a path’s contribution is proportional to δ^ℓ , where ℓ is the path length. Nevertheless, the γ , δ and η parameters allow control over the $SHCE$, so that preference can be given to a node’s bonding or bridging capabilities.

It is interesting to note the similarities between the $SHCE$ and the Katz measure, kc . Differently to many other centrality measures (such as cc and hc , where only the length of the path is important), both kc and $SHCE$ accumulate a contribution along *all* shortest paths between pairs of nodes, in proportion to δ^ℓ . However, for the Katz measure, this contribution is associated with the *source* of the path, while in the $SHCE$, we can use γ to control whether the contribution is assigned to the source, or among the intermediary nodes on the path.

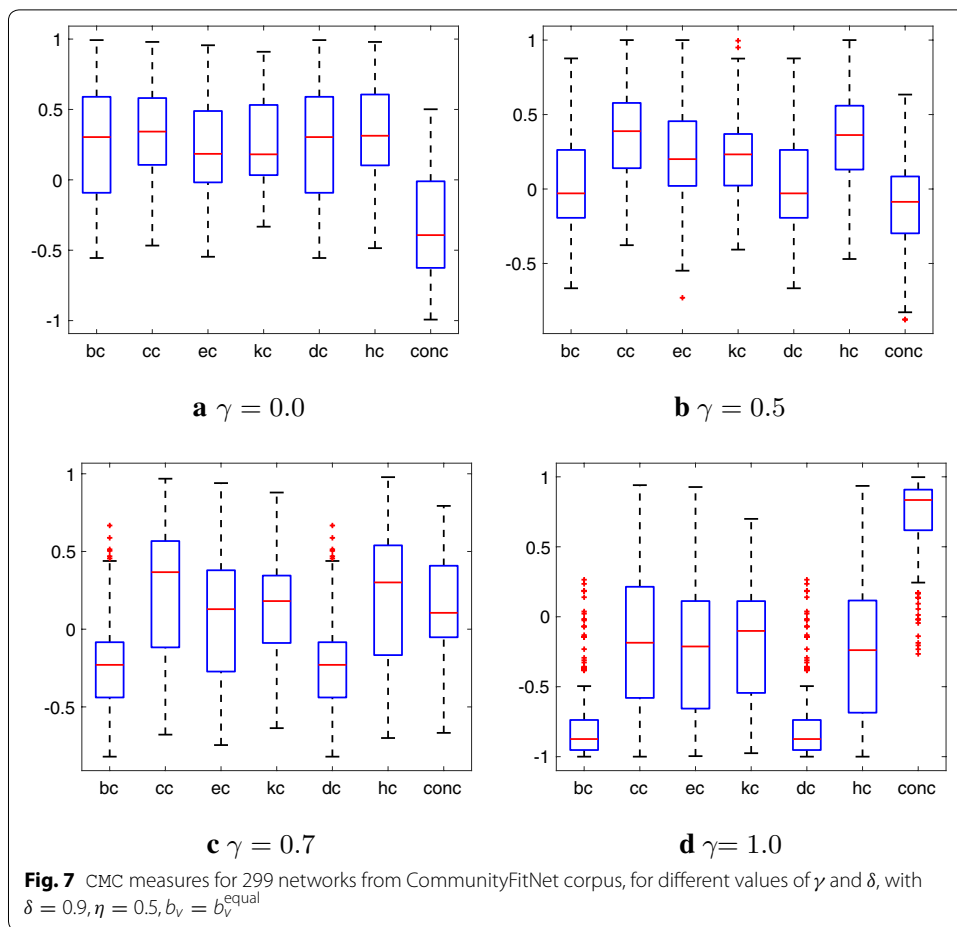
Considering the parameters of the $SHCE$, we note that δ , d_{max} , η and b_v relate to how the network is valued—the extent to which value is placed on indirect paths, and how the end-points of these paths are relatively valued. The parameter γ relates to how that value is allocated to the nodes in the network. Generally, actors bring value to the network through the paths that they occupy and that value is allocated to them proportionately, as determined by γ . The parameter η controls the value of direct connections, the more costly they are, the more value needs to attain through the indirect connections that they help form. d_{max} and δ together determine the distance horizon over which an actor can attain some value for others in the network. In the following

³ Betweenness centrality (scikit-learn) failed to compute for some networks.



analysis, we fix $d_{\max} = 10$, which for most of the involved networks exceeds or is close to their diameter and set $\delta = 0.9$.

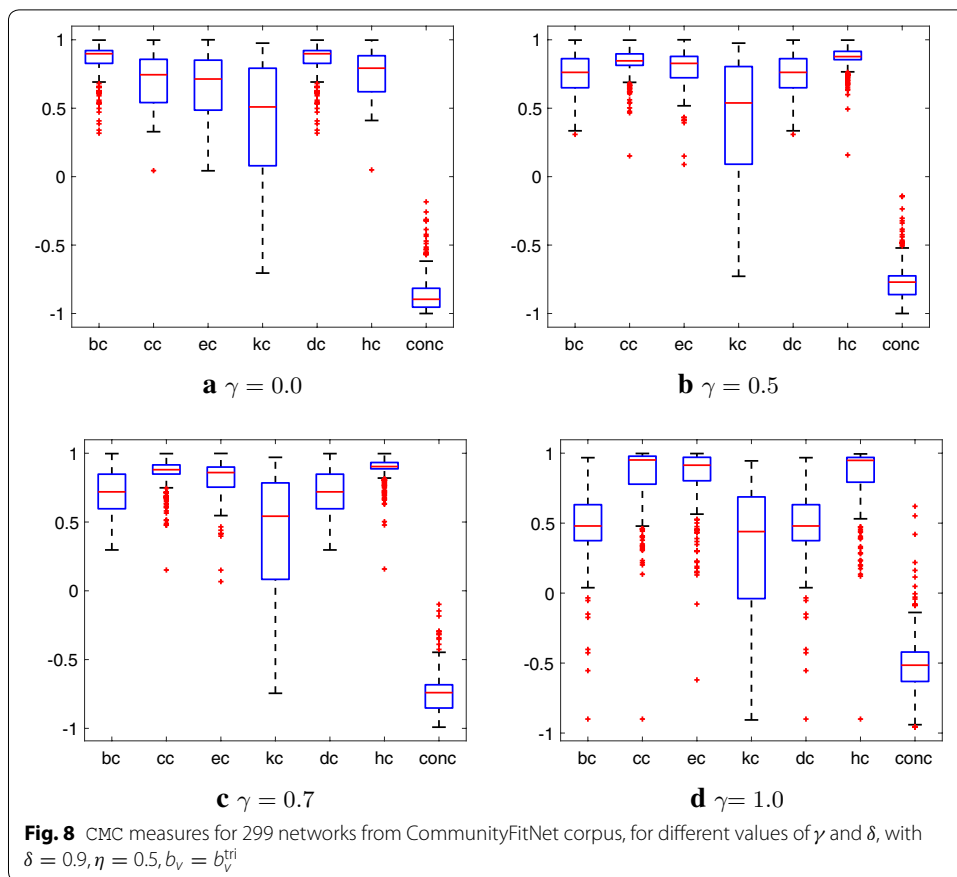
In Figs. 6 and 7, we show boxplots of the correlations of the $SHCE$ with the centralities defined in Table 3 when $\eta = 0.0$ and $\eta = 0.5$, respectively, and a constant nodal benefit function is used. Figures 8 and 9 contain the analogous boxplots for the case of the triangle benefit function. We can observe the effect of varying the γ parameter to distribute the network value in different ways. When $\gamma = 0.0$, the value from indirect links is placed fully on the intermediaries, the $SHCE$ correlates most strongly with the betweenness centrality bc and this correlation weakens as γ is raised to 1.0. At the same time, we see a strengthening of the correlation to the cc, bc and hc that value short connections from source nodes to other nodes in the network. Generally, when $\eta = 0.0$ and there is no cost associated with direct link formation, so that high degree nodes are not penalised, we see that the $SHCE$ is consistently negatively correlated with the $conc$, which values dense neighbourhoods. On the other hand, when a cost for link formation is introduced (Figs. 7 and 9), then the $SHCE$ exhibits increasing positive correlation with $conc$ as value is focused away from intermediaries. We see that the $SHCE$ becomes less well-correlated with standard centrality measures as a mixture of benefits (Figs. 7b and 9b) is valued. We also see less strong correlations with the standard centrality measures when the triangle benefit function is used. It should be



noted that, particularly, for some of the smaller networks in the dataset, these can be a high fraction of nodes that are not incident on any triangles, reducing the benefit of connecting to them to zero.

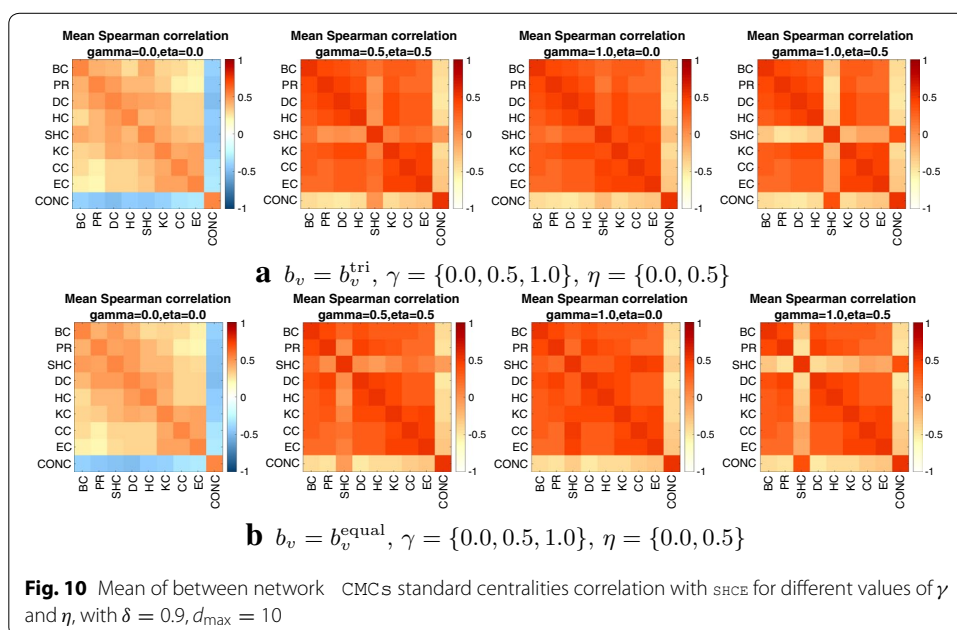
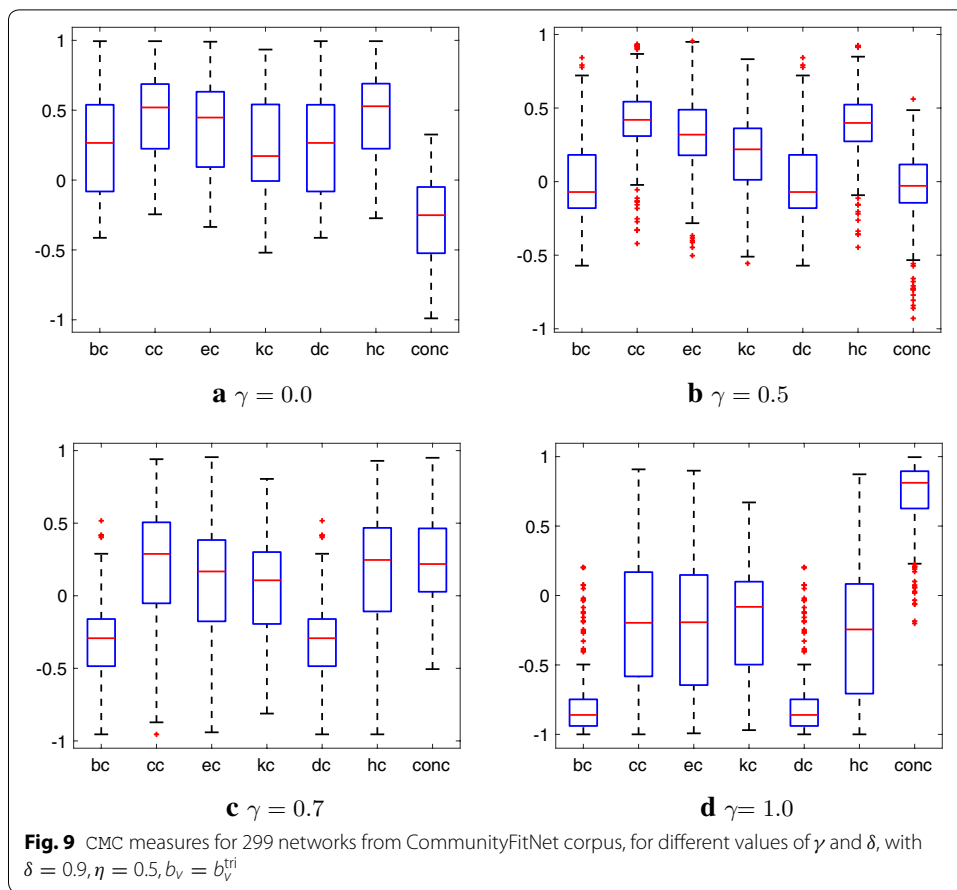
Similar to correlation analysis between centrality measures in [11], in Fig. 10, we examine the similarity of SH_{CE} with other centrality measures. Different combinations of γ, η , and δ were used to measure SH_{CE} values using both constant and triangle-based nodal benefits. The Spearman’s ρ correlation plots show that most of the pairs of centrality measures have medium-to-high positive correlation (with the exception of `conc`) with each other when compared using mean between-network CMC (the mean CMC for each pair of centrality measures across 299 networks) values. Similar to the boxplots, in these plots, for both constant and triangle-based benefits, the `conc` is negatively correlated with other measures as it values for zero values of γ and η at $\delta = 0.9$.

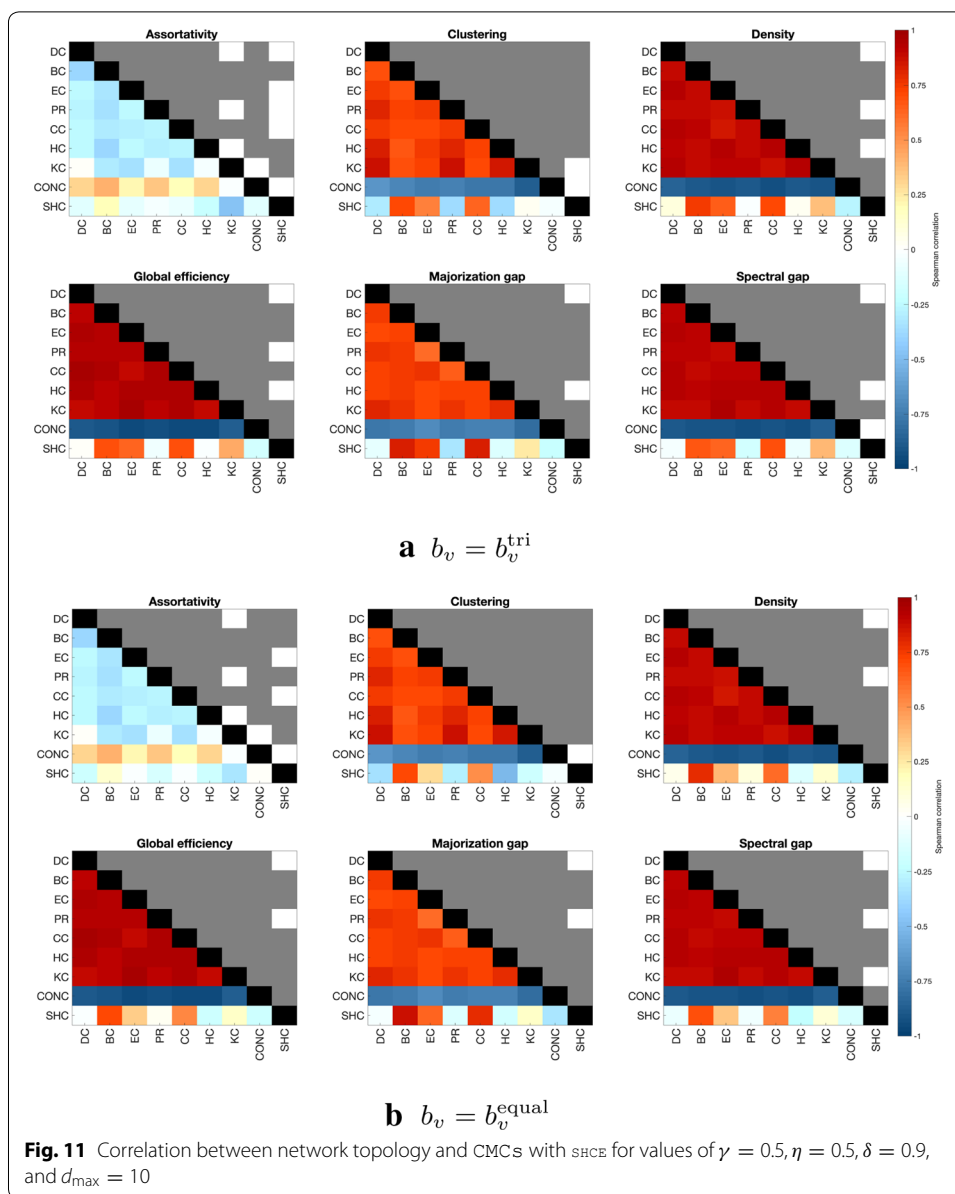
In addition to correlation between SH_{CE} and other centrality measures, we also examined the association between network properties and the CMC for different networks. We used following six out of the eight global network properties used for the similar analysis in [11]: assortivity, connection density, clustering, global efficiency, majorization gap, and spectral gap. In particular, objective of this analysis to examine



how the $_{\text{SHCE}}$ relates to the network topology as well as how it is compared relative to other centrality measures. Before results of this analysis are discussed, we briefly remind ourselves the definitions of network topological properties that were used in the analysis. Assortivity measures node’s preferences to connect with other nodes with similar degree. Clustering is the number of closed triangles in the network. The efficiency measure defined by [27] is the inverse of path connecting two nodes in the network and at global scale global efficiency is the average of efficiency for all the nodes in the network [28]. The majorization gap is the difference between empirical network and idealised threshold network [29]. It is calculated as difference in network degree sequence and its corrected conjugate sequence. Networks with high majorization gap will be distant from a threshold network and have lower CMCs [11]. Finally, the spectral gap is the difference between moduli of two largest eigenvalues of the adjacency matrix. It quantifies the extent to which a network being sparse and well connected at the same time [11].

Figure 11 shows the association between the network measures and the mean within network CMC including the $_{\text{SHCE}}$ with $\gamma = 0.5, \delta = 0.5, \eta = 0.5$ calculated for both, the triangle nodal benefit (b_v^{tri}) and the constant nodal benefit (b_v^{equal}) functions, shown in Fig. 11a and 11b, respectively. The lower triangle in each subplot indicates the Spearman correlation between CMC and the network property. The upper triangle indicates if this correlation was significant (grey) or not (while). Through our analysis





with various combinations of parameters, we observed that $SHCE$ consistently is significantly correlated with path-based network measures and negatively correlated with assortativity across various values of γ .

Overall, it may be concluded that the $SHCE$ behaves in an expected manner and aligns with other centrality measures to a greater or lesser extent, depending on the setting of its parameters. However, a single measure that allows control over a node’s bonding and bridging capabilities can be useful. For instance, an analysis of a node’s rank vs γ , can allow an analyst to better understand how the actor’s social status is composed. A low rank will indicate low status, in any case, but a rank which diminishes with γ suggests that status is being maintained mainly through bonding relationships, suggesting a route to increasing social capital would focus on enhancing its role as a bridges.

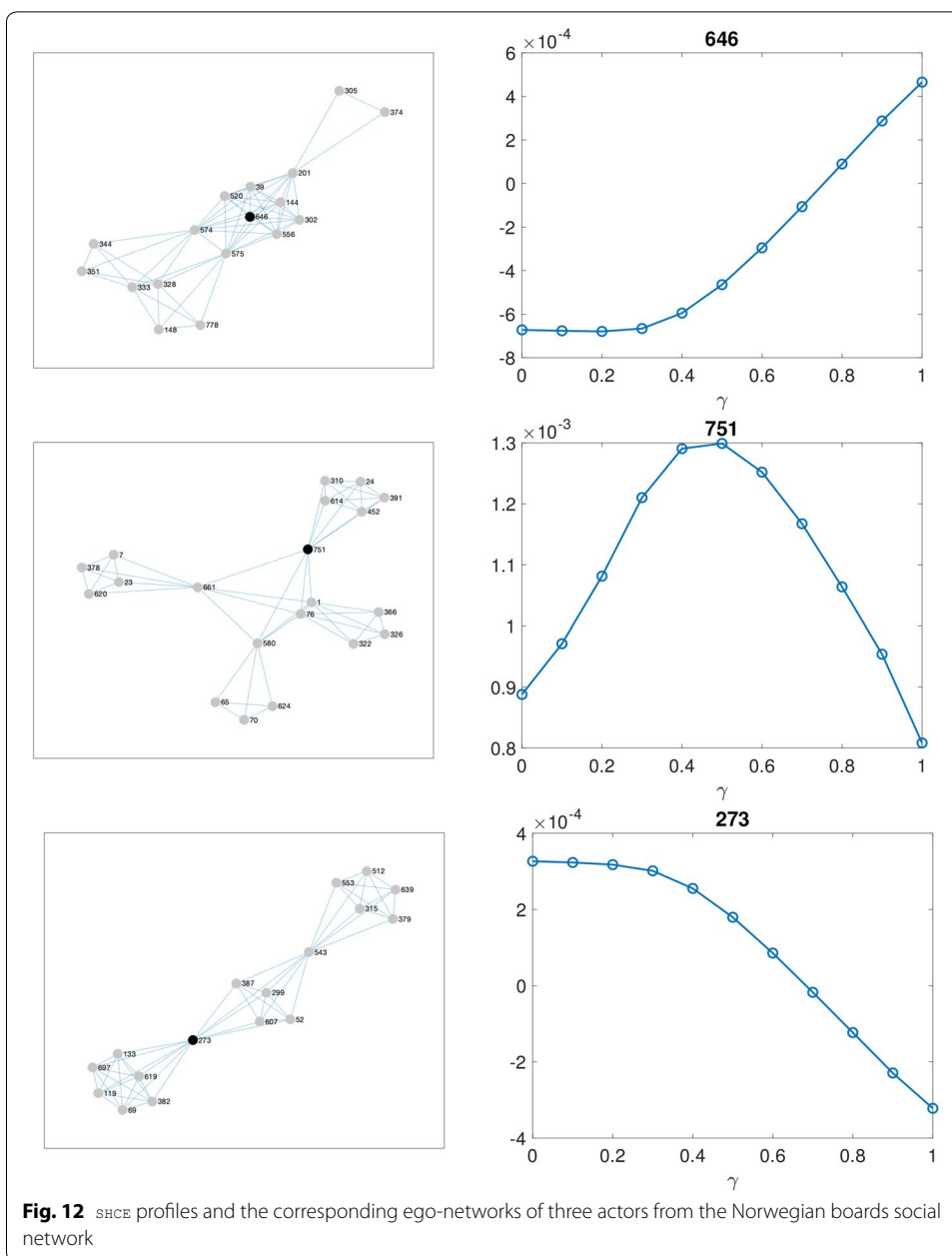
The Norwegian boards social network

We illustrate an application of the $SHCE$ in the analysis of the social network of Norwegian boards of directors introduced in [30]. This set of networks were originally used to analyse the social capital of women directors in Norway. We take the May 2011 one-mode dataset in which actors correspond to board members and a link between a pair of actors exists in the network if they are members of a common board. We extract the largest connected component of this network, which consists of 784 nodes and 2522 edges. For each actor in the network, we compute the $SHCE$ value with $d_{\max} = 10$, $\delta = 0.9$, $\eta = 0.5$, $b_v = b_v^{\text{equal}}$ and a range of γ values from 0 to 1. Thus, we allow long paths up to 10 connections to impact on the $SHCE$ and discount according to path length relatively slowly. We examine the different $SHCE$ profiles that result, where a *profile* of each actor is a graph of an actor's $SHCE$ centrality vs γ . We focus on how the profile can allow broad categories of actor to be identified. In particular, we examine at what value of γ an actor achieve their highest $SHCE$ value. A large majority of actors (83%) achieve their maximum $SHCE$ value at $\gamma = 1$, indicating that it is primarily through their bonding (over direct and indirect paths) to other actors that their social status is achieved. Only two actors, who are both female, achieve them maximum $SHCE$ score at $\gamma = 0$, indicating indicating that it is primarily through their bridging capabilities that their social status in the network is achieved. Just 6 out of 784, have a balanced profile, in which their greatest $SHCE$ value is achieved at $\gamma = 0.5$. Four out of six of these 'balanced' profiles are female. Examples of the three different profile types, are illustrated in Fig. 12, where ego-networks, extending to depth two from the ego are displayed alongside the $SHCE$ profile. We can observe in these examples, how actor 646, whose profile $SHCE$ increases with γ is bound in a tightly knit community, while actor 273 bridges along many paths between friends-of-a-friend; the balanced profile actor 751 is also a good bridge, while having many direct connections in well-connected neighbours. As another indication that female actors are somewhat more inclined to act as bridges in the social network, the actors are ordered according to the value of γ at which their $SHCE$ profile peaks, so that actors whose $SHCE$ profile peaks at $\gamma = 0$ are ordered first and those whose profile peaks at $\gamma = 1$ are last in the ordering. Focusing only on the 17% of actors whose peak is before $\gamma = 1$, in Fig. 13, we plot the cumulative proportion of females and males in that ordering. We see that females are over-represented among the low values of γ , indicating a greater tendency for female actors to get more value from the network, when that value is allocated to bridges.

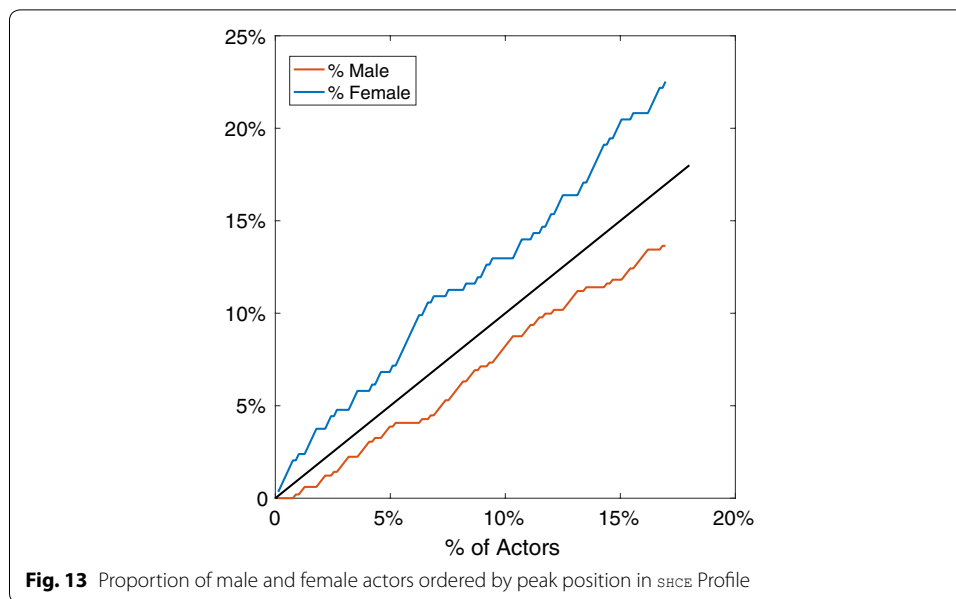
The purpose of this example is to illustrate the potential of the $SHCE$ to shed light on issues of social capital in social networks. We do not offer definitive conclusions and refer readers to [30] for a deep sociological analysis of these networks. However, we do contend that the $SHCE$ can yield deeper insights, in comparison to the betweenness centrality measure that was exploited in the original study.

Conclusion

This paper has extended the state-of-the-art on strategic network formation by proposing a new utility with associated formation game, that generalises and combines the previously proposed $CONN$ and KSH network formation games. While we have shown some



examples of efficient networks that emerge from this game, the main focus of this paper has been on a new centrality measure, that is defined as a fixed point of the linear system that spreads the benefit associated with each node in the network, among those nodes that connect to it along geodesic paths. The new centrality measure has the advantage of the Katz measure in that it depends on the connecting paths, rather than simply on path-lengths. But, more particularly, it is parameterised in a way that allows the analyst to control the way nodes are valued according to their bonding and bridging capabilities. We have benchmarked the new measure against a number of other common centrality measures and showed its application on some example networks. In future work, we



will provide a more detailed analysis of the bonding and bridging game and identify the structures that emerge as stable networks from this game.

Acknowledgements

The authors would like to thank the funding partners for supporting the research. This work was supported by Science Foundation Ireland, Grant ID is 12/RC/2289_P2.

Authors' contributions

Both authors conceived the idea and performed the literature study. NH designed the experiments, analyses and provided guidance, supervised the findings of this research, and contributed to the writing of the manuscript. Both authors read and approved the final manuscript.

Funding

This research was supported by Science Foundation Ireland, Grant ID is 12/RC/2289_P2.

Availability of data and materials

The datasets analysed in this study are publicly available online at <https://github.com/Aghasemian/CommunityFitNet>. The source code is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ IBM Ireland Limited, Dublin, Ireland. ² Insight Centre of Data Analytics, University College Dublin, Dublin, Ireland.

Received: 28 February 2020 Accepted: 3 August 2020

Published online: 17 September 2020

References

- Granovetter MS. The strength of weak ties. *Am J Sociol.* 1973;78(6):1360–80.
- Burt RS. *Structural holes: the social structure of competition.* Cambridge: Harvard University Press; 1992.
- Calvo-Armengol A, Jackson MO. The effects of social networks on employment and inequality. *Am Econ Rev.* 2004;94(3):426–54.
- Leskovec J, Adamic LA, Huberman BA. The dynamics of viral marketing. *ACM Trans Web TWEB.* 2007;1(1):5.
- Jackson MO, Yariv L. Diffusion of behavior and equilibrium properties in network games. *Am Econ Rev.* 2007;97(2):92–8.
- Seibert SE, Kraimer ML, Liden RC. A social capital theory of career success. *Acad Manag J.* 2001;44(2):219–37. <https://doi.org/10.2307/3069452>.
- Adler PS, Kwon S-w. Prospects for a new concept. *Acad Manag Rev.* 2002;27(1):17–40. <https://doi.org/10.5465/AMR.2002.5922314>.
- Putnam RD. Bowling alone: America's declining social capital. In: Crothers L, Lockhart C, editors. *Culture and politics.* New York: Palgrave Macmillan; 2000. p. 223–34. https://doi.org/10.1007/978-1-349-62397-6_12.

9. Burt R. Structural holes and good ideas. *Am J Sociol*. 2004;110(2):349–99.
10. Everett M, Valente T. Bridging, brokerage and betweenness. *Soc Netw*. 2016;44:202–8.
11. Oldham S, Fulcher B, Parkes L, Arnatkeviciute A, Suo C, Fornito A. Consistency and differences between centrality metrics across distinct classes of networks. *PLoS ONE*. 2019;14(7):e0220061. <https://doi.org/10.1371/journal.pone.0220061>.
12. Bavelas A. Communication patterns in task-oriented groups. *J Acoust Soc Am*. 1950;22:725. <https://doi.org/10.1121/1.1906679>.
13. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*. 1977;40(1):35–41. <https://doi.org/10.2307/3033543>.
14. Newman M. *Networks an introduction*. Oxford: Oxford Press; 2010. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>.
15. Page L, Brin S, Motwani R, Winograd T. The pagerank citation ranking: bringing order to the web. In: *Proceedings of the 7th international world wide web conference*. Springer; 1998. p. 161–72.
16. Everett M, Borgatti S. Induced, endogenous and exogenous centrality. *Soc Netw*. 2010;32:339–44.
17. Jackson MO, Wolinsky A. A strategic model of social and economic networks. *J Econ Theory*. 1996;71(1):44–74.
18. Jackson M. 1. In: Demange G, Wooders M, editors. *A survey of models of network formation: stability and efficiency*. Cambridge: Cambridge University Press; 2005.
19. Hummon NP. Utility and dynamic social networks. *Soc Netw*. 2000;22(3):221–49.
20. Doreian P. Actor network utilities and network evolution. *Soc Netw*. 2006;28(2):137–64.
21. Kleinberg J, Suri S, Tardos É, Wexler T. Strategic network formation with structural holes. In: *Proceedings of the 9th ACM conference on electronic commerce*. ACM; 2008. p. 284–93.
22. Narayanam R, Narahari Y. Topologies of strategically formed social networks based on a generic value function allocation rule model. *Soc Netw*. 2011;33(1):56–69.
23. Myerson R. Graphs and cooperation in games. *Math Oper Res*. 1977;2(3):225–9.
24. Jackson MO. Allocation rules for network games. *Games Econ Behav*. 2005;51(1):128–54. <https://doi.org/10.1016/j.geb.2004.04.009>.
25. Ghasemian A, Hosseinmardi H, Clauset A. Evaluating overfit and underfit in models of network community structure. *ArXiv Pre-print*; 2018.
26. Clauset A, Tucker E, Sainz M. *The Colorado Index of Complex Networks (2016)*. <https://icon.colorado.edu/>. Accessed 5 Feb 2020.
27. Latora V, Marchiori M. Efficient behavior of small-world networks. *Phys Rev Lett*. 2001;87(19):198701.
28. Ek B, VerSchneider C, Narayan DA. Global efficiency of graphs. *AKCE Int J Graphs Combinat*. 2015;12(1):1–13.
29. Schoch D, Valente TW, Brandes U. Correlations among centrality indices and a class of uniquely ranked graphs. *Soc Netw*. 2017;50:46–54.
30. Seierstad C, Opsahl T. For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in norway. *Scand J Manag*. 2011;27(1):44–54.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
