Kuikka, Vesa; Monsivais, Daniel; Kaski, Kimmo K.

Influence spreading model in analysing ego-centric social networks

Published in:
Physica A: Statistical Mechanics and its Applications

DOI:
10.1016/j.physa.2021.126524

Published: 15/02/2022

Physica A 588 (2022) 126524

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Influence spreading model in analysing ego-centric social networks

Vesa Kuikka [a,*], Daniel Monsivais [b], Kimmo K. Kaski [b,c]

[a] Finnish Defence Research Agency, Tykkikentäntie 1, PO BOX 10, 11311 Riihimäki, Finland
[b] Department of Computer Science, Aalto University School of Science, 00076, Finland
[c] The Alan Turing Institute, 96 Euston Rd, Kings Cross, London NW1 2DB, UK

ARTICLE INFO

Article history:
Received 17 December 2020
Received in revised form 23 July 2021
Available online 20 October 2021

Keywords:
Ego-centric network
Social network
Influence spreading model
Mobile phone call detail record
Centrality measure

ABSTRACT

In an earlier study one of us had developed a model of influence spreading for analysing human behaviour and interaction with others in a social network. Here we apply this model and corresponding influence centrality measures to real data of mobile phone call detail records. From this we get structures of human ego-centric networks and use a simple model, based on the number of phone calls, to describe the strengths of social relationships. To analyse 48,000 egos in their ego-centric networks we define normalised out-centrality and in-centrality influence measures, by dividing with out-degree and in-degree, respectively. With these and the betweenness centrality measures, we analyse the influence spreading in the ego-centric networks under different scenarios of link strengths between individuals reflecting the network structure being either interaction or connectivity oriented. The model reveals characteristics of social behaviour that are not obvious from the data analysis of raw empirical data or from the results of standard centrality measures. A transition is discovered in behaviour from young to older age groups for both genders and in both normalised out-centrality and in-centrality as well as betweenness centrality results.

1. Introduction

Today's digital information communication technology (ICT) with ever-growing social media allows us in a number of ways to be in contact, communicate, and interact with as well as influence other people. These lead to formation of socially connected and dynamically changing groups, communities, and societies, which can be visualised as social networks of individuals considered as nodes and interactions between them as links with weight. Due to ICT these links are not limited to friends and relatives in one's own neighbourhood, but consist of links to individuals in far-away locations and of different cultures across the globe, thus making the social network structurally complex and to some extent distance-independent. This in turn reflects as behavioural, socio-economic, political and even societal changes, as people have become networked at a larger scale, thus making spreading, exchange and sharing of information, opinion, influence, and cultural resources easier and faster. In this, the social media services and networks in particular play an increasing role, while at the same time a lot of data of the events taking place in them are generated and stored. This will allow us through data analysis and data-driven modelling to get qualitative and quantitative insight to the structure of and dynamics in social networks.

* Corresponding author.
E-mail address: vesa.kuikka@mil.fi (V. Kuikka).

https://doi.org/10.1016/j.physa.2021.126524

Large-scale social networks are composed of micro-scale network structures such as those related to relatives, close friends and co-workers, in which the individual relationships correspond to everyday social life or workplace activities. Usually, these close interactions are not dominated by processes of public opinion formation, general trends in social media or external social pressure from the outside world. Even though the content of relationships is different from public opinion or influence formation, it is likely that similar mechanisms are involved in small-scale personal networks. From this viewpoint, we develop a model of influence spreading in social networks and apply it to real ego-centric networks, based on data collected from a large mobile phone call detail records (CDRs) database by sampling egos and all the alters one link from the egos. This procedure omits interactions via individuals that are farther away from the ego. However, because this restriction is similar for the entire set of the analysed ego-centric networks, we can still make conclusions about the trends and changes in the ego-centric network structures and in the numerical values of characteristic network measures, i.e. centrality and betweenness.

Modelling of social influence is a developing field of research since the 1950's [1,2] yet new achievements have recently been attained, especially in the area of modelling influence spreading in social networks [3–10]. For more insight into this Borgatti [11] has suggested a typology of flow processes based on various trajectories that traffic may follow and the method of spread. According to this typology, social influence spreads through replication rather than transfer, in both of which trajectories can revisit network nodes. Complex contagion [7,12] is a phenomenon in social networks, in which multiple sources of reinforcement are required before an individual adopts the change of behaviour. Emotional contagion is one of the mechanisms that explains the need for multiple exposures in the spread of influence. A review of models, methods and evaluation aspects related to social influence analysis has been presented in [3].

Our approach to model ego-centric networks is that of an influence spreading model [13,14], which describes relationships in social networks. It is commonly assumed in social networks that repeated interactions both between ego and his or her alters, and between alters themselves strengthen the mutual social ties in the group. Links between individuals or nodes represent interactions and they can be bi-directional with different strengths. We assume that influence spreading is initiated from one node at a time and it propagates in the network through links. When the propagation via multiple paths leads to the same target node, we use the result from the probability theory describing mutually non-exclusive events. The same probabilistic method is used to combine different paths starting from a source node with common partial paths before branching of the paths. We assume that propagation after branching via different paths is independent and the states of the nodes have no effect on the spreading probabilities. This is an approximation for describing influence, opinion or behaviour spreading.

In this paper, we propose an influence centrality measure based on the influence spreading model. As a result of this, the methods for analysing spreading processes on a network structure and definitions of influence metrics are consistent with each other. A new form of betweenness centrality measure [2] is also proposed based on calculating the change of the influence centrality measure when one node, at a time, is removed from the network structure. Nodes with high betweenness [2] may have considerable influence within a network because of their role to control over communication between others.

2. Influence spreading model

In Network Science there is a variety of mathematical measures of centrality that focus on different concepts and definitions of what it means to be important or central in the network. Nodes can be centrally located or peripheral in a network structure and they can have different roles, i.e. serve as mediators between different parts of the network or operate as efficient spreaders. Definitions of eight commonly used centrality measures have been presented in a recent review article [15] and they can be categorised in centrality indexes [1,15,16] and influence metrics [8–10], but most of the definitions in the literature have features of both of these aspects. Of these two categories the latter – the influence metrics – offers a socially relevant perspective to centrality as it quantifies the influence or role of each individual in a social network.

The simplest centrality measure is the degree, which is the number of links that are incident to the node. In directed networks, one defines the in-degree and out-degree as the number of incoming and outgoing links, respectively. Examples of other measures are closeness centrality, betweenness centrality, eigenvector centrality, Katz centrality and random walk centrality, see [15]. The definition of centrality in this paper is not equivalent to any of these definitions. However, it has some common features with the Katz centrality [1] as it generalises the degree centrality and the closeness centrality by taking into account not only the immediate neighbours or not only the shortest paths from a node to other nodes.

Here we present a new methodology [13,14] to model influence spreading in a network of people, where the nodes describe individuals and the links describe connections between them. In social networks there are typically, multiple paths between a pair of nodes, with possible common parts. Here we devise a mathematical model that considers all paths between all the node pairs of a network. In our model, nodes can appear multiple times in a spreading process. We assume that influence spreading and opinion formation occur as a gradual development with loops in the paths describing recurrent events. From this information we construct an influence spreading matrix, or the probability matrix, that describes influence spreading from source nodes to all other nodes in the network. Because of the complex structure of social networks, the spreading probability from node A to node B is not equal to the probability from node B to node A. One consequence of this is that peripheral nodes that are locally densely connected can have a considerable effect on

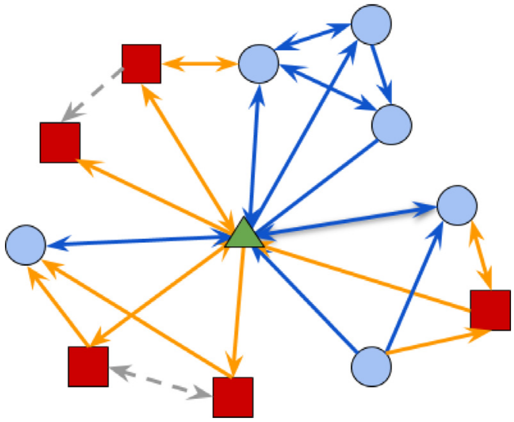

Fig. 1. An example of directed ego-centric mobile phone based social network. The ego is the focal node represented by the green triangle, while alters who are subscribers of the service provider are represented by blue circles and non-subscribers by red squares. Links between subscribers are in blue, between a subscriber and a non-subscriber in yellow, and possible (but not detected) links between non-subscribers are represented by dashed grey lines. Unidirectional arrow indicates a singly directed link from a node to another node i.e. $k \rightarrow l$ to and bidirectional arrow indicate two separate directed links between the pair of nodes $k \rightarrow l$ and $k \leftarrow l$.

other parts of the network. Thus the influence spreading matrix is not symmetric, except in the rare case of symmetric network topology.

In the process of forwarding news and information, recurrent visits on a node as well as the assumption of state independence of nodes may not be so good approximation. We assume that these kinds of events do not have a major effect on social relationships in the ego-centric networks. After all, if the amount of information transmission between two nodes is correlated with the volume of more abstract content of opinion exchange, this would not have a significant effect on the results of the influence spreading model. The ratio of social exchange to information forwarding may be different, for example, between members of a family and between co-workers.

Here, we give an idea of how the model can be programmed with a computing language as a more detailed pseudo-algorithm has been presented in [13]. The algorithm considers all possible paths from a source node to a target node. For every path, the probability of propagation is calculated by multiplying all the link weights along the path, where each link weight w_{kl} from a node k to node l is interpreted as the probability of propagation via the link. Let us denote a path i with $\mathcal{L}(i)$ in the list of paths $\mathcal{L}$, then $W_{\mathcal{L}(i)}$ is the product of link weights w_{kl} on that path. The maximum length of paths when computing the probabilities can be limited by the model parameter L_{max}. Using this parameter is necessary because we allow paths with loops. Also, we assume that the spreading process is state-independent, meaning that the propagation of influence coming via multiple paths to a node is passed repeatedly over to its neighbouring nodes. Allowing loops and assuming state-independent propagation are consistent properties of a model that describes social influence. A model with self-avoiding paths and state-dependent propagation of events could describe formal information transmission when nodes have memory.

Next, we write an iterative formula for calculating the spreading probabilities in the network by considering paths between a source node and a target node. Paths from source node s to target node t are combined pairwise in the descending order of the lengths of the common path before the first branching of the two paths. Two paths $\mathcal{L}(i_1)$ and $\mathcal{L}(i_2)$ are combined into the path $\mathcal{L}(i)$ for which the probability value $P_{\mathcal{L}(i)}$ reads as follows:

$$P_{\mathcal{L}(i)} = P_{\mathcal{L}(i_1)} + P_{\mathcal{L}(i_1)} - \frac{P_{\mathcal{L}(i_1)} P_{\mathcal{L}(i_2)}}{W_{\mathcal{L}(i_1) \cap \mathcal{L}(i_2)}}. \qquad (1)$$

In this procedure, the order of processing the paths is a crucial point: Indexes i_1 and i_2 are selected in the descending order of path length of the common path part of the two paths $\mathcal{L}(i_1)$ and $\mathcal{L}(i_2)$, denoted here with $\mathcal{L}(i_1) \cap \mathcal{L}(i_2)$. All the original paths in the list are processed only once and combined paths are processed in the same way as the original paths in the network. In the procedure, the processed paths are removed and the combined paths are inserted into the list $\mathcal{L}$, i.e. the combined path $\mathcal{L}(i)$ is inserted and the paths $\mathcal{L}(i_1)$ and $\mathcal{L}(i_2)$ are removed. The procedure starts with the two paths having the longest common path length L of path $\mathcal{L}(i'_1) \cap \mathcal{L}(i'_2)$ with $P_{\mathcal{L}(i'_1)} = W_{\mathcal{L}(i'_1)}$ and $P_{\mathcal{L}(i'_2)} = W_{\mathcal{L}(i'_2)}$, where $W_{\mathcal{L}(i'_1)}$, $W_{\mathcal{L}(i'_2)}$ and $W_{\mathcal{L}(i'_1) \cap \mathcal{L}(i'_2)}$ are the products of all link weights along paths $\mathcal{L}(i'_1)$, $\mathcal{L}(i'_2)$, and $\mathcal{L}(i'_1) \cap \mathcal{L}(i'_2)$, respectively. If there are more than two paths with the same common path length, these paths can be processed in any order. A numerical example in [13] illustrates the algorithm in practice. The probability of influence spreading between the two nodes is the final result of the algorithm after all paths from the source node s and the target node t have been processed:

$$C(s, t) = P_{\mathcal{L}(i)}. \qquad (2)$$

Eq. (2) defines the matrix element $C(s, t)$ of the influence spreading matrix.

The non-normalised out-centrality and in-centrality measures for nodes s and t in a network of N nodes can be defined as follows

$$C(s)^{(out)} = \sum_{t=1}^{N} C(s,t). \tag{3}$$

$$C(t)^{(in)} = \sum_{s=1}^{N} C(s,t). \tag{4}$$

The normalised versions can be obtained by dividing these expressions by N or $N-1$ depending on whether the diagonal elements of the influence matrix C are set to one or zero. Then the corresponding betweenness centrality measure for the node m can be defined as

$$b_m = \frac{C - B_m}{C}, \tag{5}$$

where

$$C = \sum_{s,t=1}^{N} C(s,t), \tag{6}$$

and B_m is calculated similarly to C with the node m removed from the network structure G as

$$B_n = \sum_{\substack{s,t=1 \\ m \notin G}}^{N} C(s,t). \tag{7}$$

The classical closeness and betweenness measures (see the review article [15]) are based on the number of shortest paths between two nodes in a network. Definitions in Eqs. (3)–(5) are generalisations that take into account more detailed structure of the network.

In order to complete the definition of the model, an expression for each link weight w_{kl} from node k to node l needs to be assigned. For this, we propose the following functional form:

$$w_{kl} = 1 - (1-\alpha)^{n_{kl}}, \tag{8}$$

where n_{kl} is the number of interactions from node k to node l, and $0 \leq \alpha \leq 1.0$ is a tuning parameter. Eq. (8) is based on the formula of mutually non-exclusive events from the probability theory consistently with Eq. (1). If an empirical expression were available for the relationship between the link weight and the number of interactions then it could be used instead of Eq. (8). The parameter α determines two different classes (or facets) of the influence spreading processes through the network. The first class, we called it a "connectivity" facet, is expressed in the range $\alpha \approx 1$. In this regime, the term $(1-\alpha)^{n_{kl}}$ is considerably small after the exponentiation, and therefore the link weight factor w_{kl} is closer to 1. In this case, our definitions for centrality and betweenness behave like the classic definitions, such that the connectivity of the network plays a central role in the influence spreading process. On the other hand, in the range $\alpha \approx 0$, the number of pairwise interactions n_{kl} plays an important role, as the w_{kl} depends strongly on n_{kl}. In such a regime, the model expresses a second class, i.e. an "interaction" facet, where the centrality measures represent the influence that nodes have in terms of their pairwise interactions in the ego-centric network.

The influence centrality measures of Eqs. (3) and (4) and the general definition of degree centrality are related in the way that the degree centrality is a special case of the influence centrality when $\alpha = 1$ and the maximum path length in computing the probabilities of the model is limited to $L_{max} = 1$. The statement holds for in-degree, out-degree and general (undirected) degree. For the general degree the condition for $L_{max} = 1$ is not necessary because the ego-centric network is defined in such a way that there is always a link between an ego and an alter (see Fig. 1). The influence centrality measure, with $\alpha = 1$ and $L_{max} = 1$, simply counts the number of alters, which is the same as the value of the ego's degree in the ego-centric network.

When comparing the definition of out-degree with influence out-centrality of Eq. (3) (and the definition of in-degree with influence in-centrality of Eq. (4) for $\alpha = 1$ but without limiting the value of L_{max}, we can see from Fig. 1 that influence out-centrality (and influence in-centrality) includes nodes that are not included in the definition of out-degree (in-degree). This holds more generally when $0 < \alpha < 1$ as pairwise interactions of alters are taken into account. Interactions between alters have a high effect on influence centrality when the ego and alters have strong connections, and strong interactions between alters have only a minor effect if the ego and the alters have weak connections.

3. Data

The dataset analysed in this research work consisted of Call Details Records (CDRs) of a mobile phone service provider from a European country, the name of which cannot be disclosed. The dataset includes information about all the communication events of the subscribers of the mobile phone service. Each subscriber is described by a unique identifier,

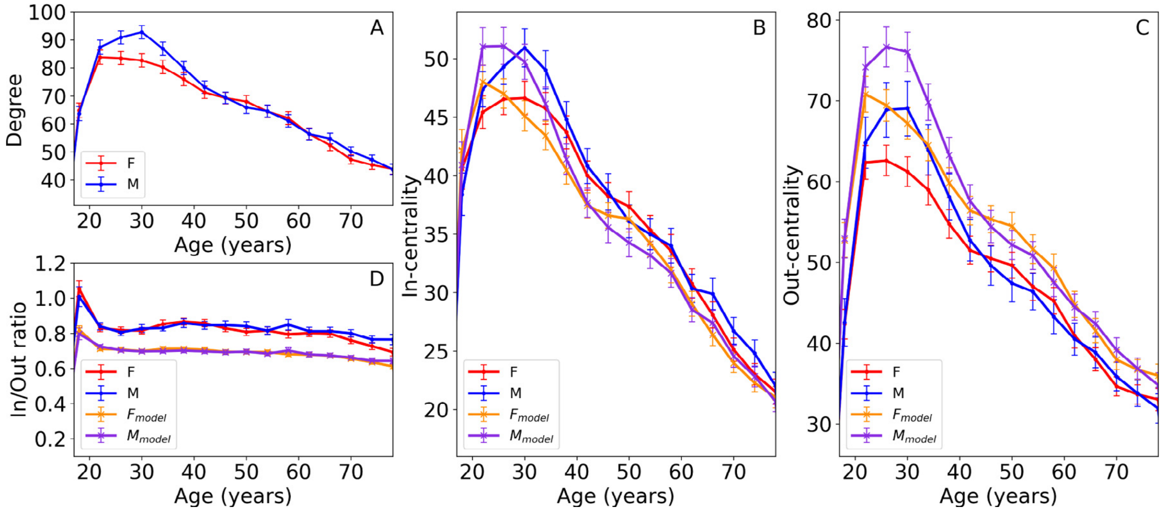

Fig. 2. The degree and influence centrality measures as a function of the ego's age: average degree (panel A), in-degree centrality and influence in-centrality (panel B), out-degree centrality and influence out-centrality (panel C) and In/Out ratios (panel D; i.e. in-degree centrality/out-degree centrality ratio and influence in-centrality/influence out-centrality ratio) for males (blue lines) and females (red lines). These curves are compared to the influence centrality of the influence spreading model, for males (violet) and females (orange), when the link weights of each ego-centric network were set according to Eq. (8) and $\alpha = 0.5$. Based on statistical analysis the error bars were added to indicate at which age the female values differ statistically significantly from the male values.

which was generated by the service provider before handling the dataset, such that the identity of each subscriber remains anonymous. The dataset includes demographic information of the subscribers such as the age, gender, location of the most used communication tower, postal code, and the starting and ending date (if so) of the service contract. In this work, we use a sample of a bigger dataset, including more than 3 billion events (calls and text messages) of 10 million subscribers of the service provider in 2007. The dataset contains all the communication events of each of its subscribers, therefore it includes communication events with phones that are not subscribers of the service provider (landlines and subscribers of other providers) such that the number of different identifiers in the dataset is more than 50 million.

Each CDR item contains the following information:

- Identifier of the person (anonymised) starting the communication event ("caller")
- Identifier of the person (anonymised) receiving the communication event ("callee")
- Duration of the communication event (0 in case of text message)
- Type of the communication event (call or text message)
- Time and date when the communication event occurred

From the original dataset, we chose a sample of subscribers from those whose age, gender, and postal code were provided (for some cases in the original dataset this information was missing). For each age group in the age interval 18–77 years and for each gender (male and female), 200 subscribers were randomly chosen, such that the final sample included $60 \times 2 \times 200 = 48{,}000$ individuals.

For each subscriber (or ego) we built the contact network (or ego-centric network), where the ego represents a focal node, connected by a link with each person (or alter) with whom the ego had at least one communication event within the studied period of time. We take into account the directionality, thus there is a directed link between the ego and an alter (or vice-versa), if there was a communication event started by the former to the latter. In addition, we search for connections between the alters of the ego, such that if there were communication events between two alters, a link joining them was added to the ego-centric network. Here, it is important to note that, as the dataset includes CDRs either between subscribers, or between a subscriber and non-subscriber, it is not possible to know if there were communication events between two non-subscribers, and because of this our ego-centric networks will not include this kind of links. In Fig. 1 an example of an ego-centric network is shown.

We generated an ego-centric network for each of the 48,000 egos, and then for each network, we calculate the number of communication events that occurred in all of its directed links during the whole period of time, studied. For the link joining the node k and l, we assigned the total number of communication events, the node k made to the node l, to the variable $n_{kl} \geq 0$, which is used in the definition of link weight within the model.

4. Results

Using the conventional definition of the degree and degree centrality measures, we calculate them as averages over the sampled egos, grouped in age and gender cohorts. In Fig. 2 we depict the total degree as well as the in-degree and out-degree centrality measures for each age and gender cohorts. Here it can be seen that the degree and degree centrality measures have a strong dependence on age, such that they reach maximum values for the age cohort of around 30-years old, and then they show a monotonous decrease as the age of the cohort increases. This observation is consistent with similar findings reported in the literature [17–19], and it indicates that in terms of the connectivity younger people have more central roles in the society. Also as for the gender difference, only in the age range at which the maximum degree is reached (i.e. 26–34 year-old cohorts), the degree and degree centrality measures for males are significantly higher than for females. Outside this age range, males and females have similar degree and degree centrality values in each of the remaining age groups. It can also be seen that for both genders and regardless of the age group, the in-degree centrality tends to be smaller than the out-degree centrality, as illustrated by the ratio of these two quantities, presented as the In/Out-ratio in the right-most panel of Fig. 2. The only exception is for the 19 year-old age cohort, where this tendency, is slightly reversed, i.e. the In/Out-ratio ≥ 1. Furthermore, it can be seen that for older age cohorts the In/Out-ratios for males and females behave similarly up to about 65 year-old, after which the males seem to have higher values than females.

In order to determine the influence centrality measures of each sampled ego we use the influence spreading model, in Section 2, and Eq. (8) to define the link weights w_{kl} of the ego-centric networks. For this, we introduce the parameter α, used as a base of an exponential function with n_{kl}, i.e. the total number of communication events from the node k to the node l, as the exponent. Here the link weights of the model, w_{kl}, are evaluated for following values of the parameter $\alpha = 0.01$, 0.025, 0.05, 0.1 and 0.5. For all the calculations, we set the parameters required by the model as follows: $L_{max} = 10$ and kept them fixed. This model is then used to measure the influence centrality of the nodes of the ego-centric networks, thus including the effects of the networks where the nodes are embedded. In Fig. 2 we show the results obtain when α is set to 0.5, such that the networks are considered to be in the connectivity facet. The overall trends of the influence centrality measures defined by the model are similar with those obtained by using the usual definitions of degree centrality measures, differing only by a scaling factor, but preserving the overall shapes of the curves. Hence as in the case in-degree and out-degree centrality the influence in-centrality and influence out-centrality as a function of the ego's age and for both genders reach maximum values in the age range of 24–32 years old beyond which they decrease almost monotonically for older age groups.

In the connectivity facet (i.e. when α is large or $\approx$1), the males in general have larger influence in-centrality and out-centrality values than females, in the younger age groups up to 40 year-old at which point there seems to be a slight crossover with very small difference between the curves, as can be seen in the second and third panels of Fig. 2. In the case of the interaction facet (i.e. when α is small or $\approx$0) we observed the behaviour of the influence in-centrality and out-centrality curves to be similar for all ego's age cohorts and for both males and females with those in the connectivity facet.

Next we demonstrate how a new metric can be defined by combining two different measures, i.e. the degree and the influence centrality presented before, to highlight specific characteristics of influence spreading. The in-degree and out-degree centrality measures in Fig. 2 depict the general trend of degree centrality, but there are differences between the influence in-centrality and influence out-centrality results when the parameter α of the link weight model in Eq. (8) is varied. Since varying the parameter α induces different scaling for the in-centrality and out-centrality measures, we scale the influence in-centrality and out-centrality with the in-degree and out-degree, respectively, to get the results for the normalised centrality measures, presented in Fig. 3. These normalised quantities show very different features depending on the influence facet we observe, i.e. for different values of α. For the normalised in-centrality, in the connectivity facet (i.e. when α is large or $\approx$1) the younger people have larger values than the other age cohorts, showing a maximum for the youngest, i.e. about 20 year-old age cohort then decreasing quite rapidly to about 40 year-old age cohort to stay at that fairly low levels but higher for females than for males. In the interaction facet (i.e. when α is small or $\approx$0), the normalised in-centrality behaves quite similarly with the exception that from the minimum of around 40 year-old age cohort its values increase gradually, once again with female at higher level than males. The normalised out-centrality, in the connectivity facet, shows a peak at about 20 year-old age cohort but comes down to about 30 year-old age cohort and stays at level up to the oldest age cohorts both for females and males indistinguishably. In the interaction facet the normalised out-centrality peaks nearer to 30 year-old age cohorts, then decreasing to about 40 year-old age cohort and plateauing to about 60 year-old age cohort and after that decreasing again with females at a higher level than men.

In addition we have determined the betweenness centrality as a function of the ego's age and for both genders, as averages over the 48,000 ego-centric networks for 5 different values of the parameter $\alpha = 0.01$, 0.025, 0.05, 0.1, and 0.5, i.e. moving from the interaction facet ($\alpha \approx 0$) towards the connection facet ($\alpha \approx 1$). In Fig. 4 we show the results. In the interaction facet, females have considerably larger betweenness centrality than males, whilst in the connectivity facet, there are no significant gender differences. We also see that the betweenness centrality for $\alpha = 0.01$ and 0.025 again shows quite prominent peaks at about 30 year-old age cohort for both genders but for females with higher values than for males throughout all ages. For $\alpha = 0.05$ and larger, the peaks get less prominent and tend to disappear, and the difference between females and males dissolves and gets slightly reversed for 40-years and older age cohorts.

5. Discussion

In this study, we have focused on analysing human behaviour in mobile phone based ego-centric social networks using an influence spreading model. In line with the earlier studies [17,18] we found that the average degree and the influence

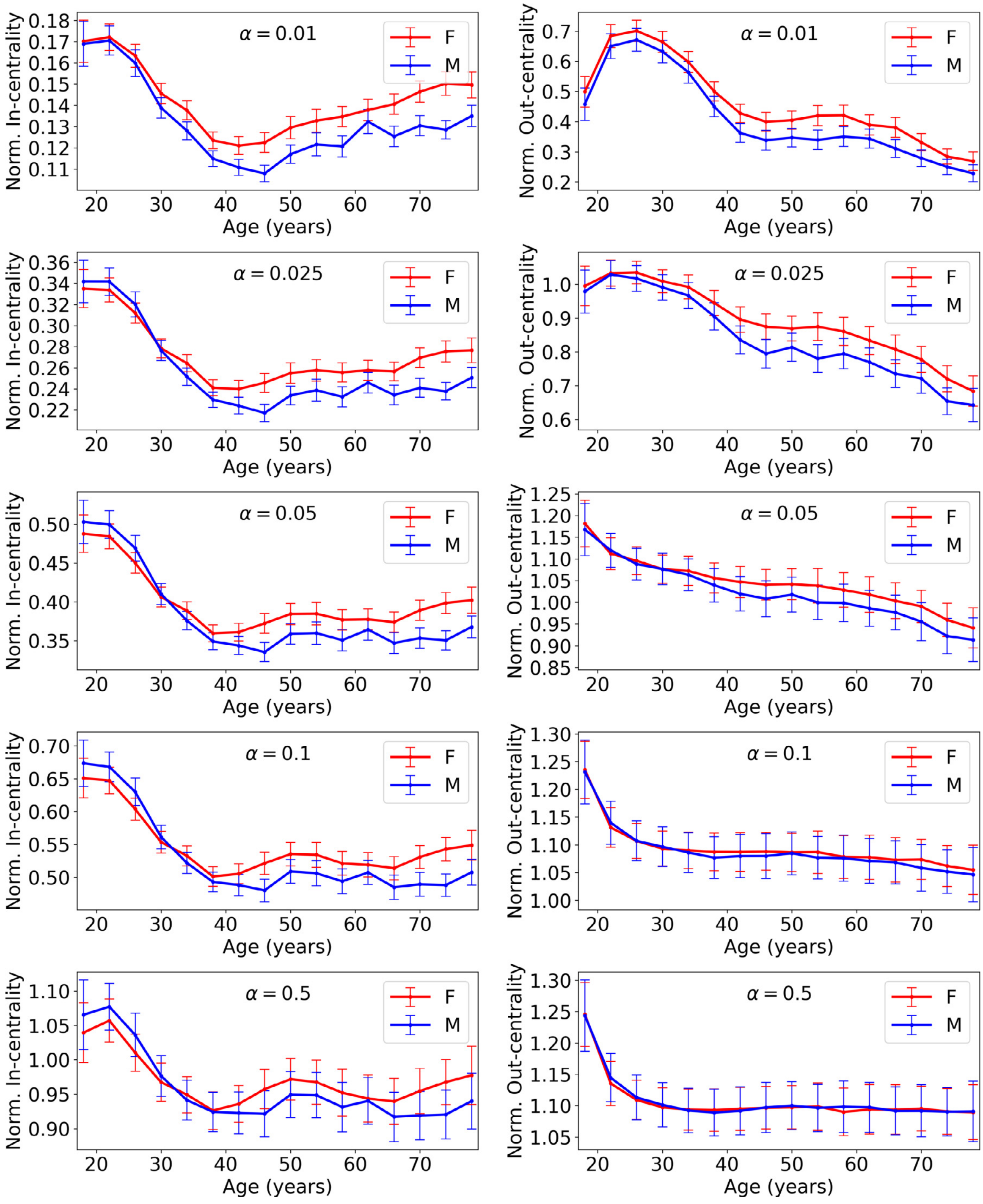

Fig. 3. The normalised influence in-centrality (on the left) and out-centrality (on the right) for males (blue) and females (red) as a function of the ego's age and for five different social influence facets of the link weight model (Eq. (8)) parameter $\alpha = 0.01$ (top row), 0.025 (2nd row), 0.05 (3rd row), 0.1 (4th row), and 0.5 (bottom row). Error bars indicate at which age the female values differ statistically significantly from the male values.

centralities (influence in-centrality and out-centrality) for males and females peak around 30 year-old age cohort, after which they both decrease monotonically (Fig. 2). This reinforces the notion that younger people tend to be more socially active by devoting more time and effort to generate and keep connections within the society they are immersed, whilst older people tend to keep less but more meaningful connections. Moreover, in agreement with the findings in [18] young

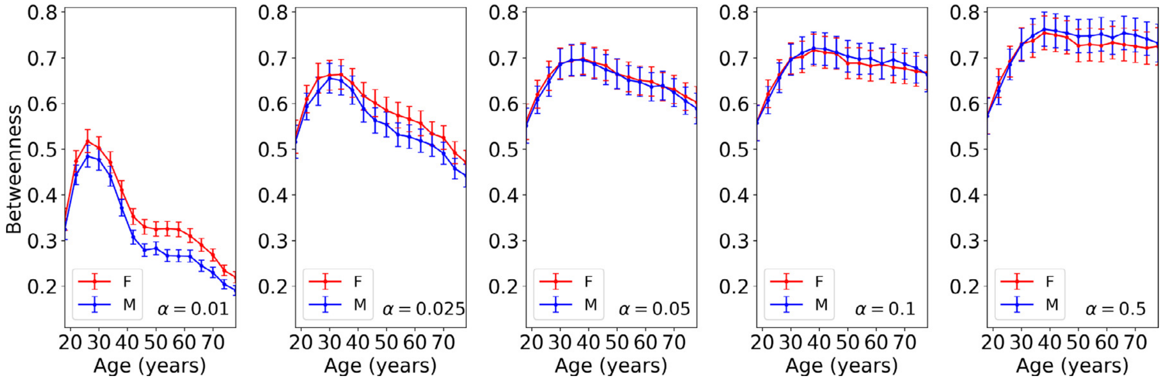

Fig. 4. The ego's betweenness centrality as a function of the ego's age for males (blue) and females (red). The averages are calculated from the 48000 ego-centric networks, for five different values of the link weight parameter $\alpha = 0.01$, 0.025, 0.05, 0.1, and 0.5, shown in each panel. The statistical analysis based error bars indicates that in the interaction facet ($\alpha \approx 0$), females have considerably larger betweenness than males, whilst in the connectivity facet there are no significant gender differences.

male cohorts have slightly larger influence centrality values than females of the same age, but there is a crossover at around the age of 40 years after which the values for females seem to be slightly higher.

As for the In/Out ratios (i.e. in-degree centrality/out-degree centrality and influence in-centrality/influence out-centrality ratios in the last panel of Fig. 2) for both males and females peak around the 19 year-old age cohort after which they decay reaching minima for the oldest age cohorts. There is no significant difference between the male and female curves for all the age cohorts. This ratio can be interpreted as the relative tendencies of the individuals to interact with their connections directionally either by receiving calls or initialling them. In general it is known that older people tend to reduce their interactions within their ego-centric networks, which is reflected as the decay of the In/Out ratios due to older people receiving cumulatively fewer calls than they initiate. Towards the end of the curves for over 60 year-old age cohorts the female curve decays slightly (or more visibly in the usual degree centrality measures) faster than the male curve, which due to their heightened calling activity to their own children especially daughters could be attributed to the grand-mothering effect [20]. The peak values in the In/Out ratios for both genders of the 19 year-old age cohort reflects strong social connectivity within the age cohort with an increased tendency to receive calls.

The normalised influence centrality based on our influence spreading model represents a novel measure of the influence that individuals experience from and exert to their social network, unveils interesting and different dynamics depending on the influence facet applied. One observation that is consistent with the In/Out ratios in Fig. 2, is that in the interaction facet normalised in-centrality is increasing and normalised out-centrality is decreasing for older than 40 year-old groups for both genders. These trends turned out to be much weaker or non-existing in the connectivity facet ($\alpha \approx 1$). Our interpretation is that for the older age groups there is a shift in the interaction based social behaviour but not in the connection based behaviour.

The youngest age groups under 30 year-old are both more influential in their social network and more influenced by their social network when compared to the older cohorts. This behaviour holds in the interaction facet ($\alpha \approx 0$) and in the connectivity facet ($\alpha \approx 1$). In the interaction facet, a maximum around the 30 year-old age group is shown particularly in the normalised out-centrality, while younger age groups have as high normalised in-centrality as 30 year-old. The normalised out-centrality in the connectivity facet shows a clear drop around the 20 year-old age group and for older than 30 year-old egos the normalised out-centrality is almost a constant. The normalised out-centrality in the connectivity facet behave similarly for both genders. In the interaction facet, female egos have both larger normalised in-centrality and out-centrality values but in the connectivity facet young male cohorts under 40 year-old have larger normalised in-centrality values than females of the same age.

In order to see how different facets of influence spreading showed age and gender differences in the betweenness centrality measure, we varied the parameter α, i.e. the link weights between the egos and alters. In the interaction facet (i.e. small α, leftmost panel in Fig. 4), the cohorts in the middle adulthood tend to be more relevant for the influence spreading process than other age cohorts, such that females have a larger influential role than males throughout all ages. It is also seen that the youngest age cohorts have a somewhat smaller influential role in the spreading process, which could be understood from the fact that these cohorts have recently entered the economic/labour part of the society and they are still in the process of generating a relevant role in it. Furthermore, in the interaction facet it is seen that 55–65 year-old females showed sizeably larger betweenness centrality than the similar age males have. This could be attributed to the fundamental role that females at that age (usually in the grand-parenting stage) play as joining and supporting bonds in the extended family. In the connectivity facet (i.e. large α, rightmost panel in Fig. 4), the importance of youngest cohorts as bridging elements of influence spreading, is the least prominent of all the age cohorts. The betweenness centrality grows gradually for the subsequent age cohorts until the age of 40-year, after which the relevance of each group is practically the same, though a bit larger for males than females. This observation, in terms of the connectivity facet of the influence spreading process, could show that the role of individuals play in their local networks as bridges for influence spreading, is established in the first stages of adulthood, after which a well-established relevance is preserved for the subsequent ages.

CRediT authorship contribution statement

Vesa Kuikka: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Daniel Monsivais:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Kimmo K. Kaski:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Daniel Monsivais and Kimmo Kaski acknowledge support from the EU Horizon2020 project SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytcss rant agreement ID: 871042.

Acknowledgements

We would like to thank Dr. Kunal Bhattacharya for useful discussions and suggestions in the early stages of this research. DM and KK acknowledge support from the EU Horizon2020 project SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics, Grant agreement ID: 871042.

References

[1] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1953) 39–43.
[2] L. Freeman, A set of measures based on betweenness, Sociometry (1977) 35–41.
[3] K. Li, H. Zhang, H. Huang, Social influence analysis: models, methods, and evaluation, Engineering 4 (1) (2018).
[4] D. Kempe, J. Kleinberg, E. Tardeos, Maximizing the spread of influence through a social network, Theory Comput. 11 (4) (2015) 105–147.
[5] A. Kirkley, G.T. Cantwell, M.E.J. Newman, Belief propagation for networks with loops, Sci. Adv. 7 (17) (2021).
[6] B. Zhou, S. Pei, L. Muchnik, X. Meng, X. Xu, A. Sela, S. Havlin, H.E. Stanley, Realistic modelling of information spread using peer-to-peer diffusion patterns, Nat. Hum. Behav. 4 (2020) 1198–1207.
[7] D. Centola, How Behavior Spreads, the Science of Complex Contagions, Princeton University Press, 2018.
[8] G. Lawyer, Understanding the influence of all nodes in a network, Sci. Rep. 5 (8665) (2015).
[9] B.A.N. Travençolo, L. da F. Costa, Accessibility in complex networks, Phys. Lett. A 373 (1) (2008) 89–95.
[10] F. Bauer, J.T. Lizier, Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach, Europhys. Lett. 99 (6) (2012).
[11] S.P. Borgatti, Centrality and network flow, Social Networks 27 (1) (2005) 55–71.
[12] D. Centola, M. Macy, Complex contagions and the Weakness of long ties, Am. J. Sociol. 113 (3) (2007).
[13] V. Kuikka, Influence spreading model used to analyse social networks and detect sub-communities, Comput. Soc. Netw. 5 (1) (2018) 12.
[14] V. Kuikka, Modelling community structure and temporal spreading on complex networks, Comput. Soc. Netw. 8 (1) (2021) 13.
[15] S. Gómez, Centrality in networks: finding the most important nodes, in: P. Moscato, N.J. de Vries (Eds.), Business and Consumer Analytics: New Ideas. Part III, 2019, pp. 401–434, (Chapter 8).
[16] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, Social Networks 32 (2010) 245–251.
[17] C. Wrzus, M. Hänel, J. Wagner, F.J. Neyer, Social network changes and life events across the life span: A meta-analysis., Psychol. Bull. 139 (1) (2013) 53.
[18] K. Bhattacharya, A. Ghosh, D. Monsivais, R.I. Dunbar, K. Kaski, Sex differences in social focus across the life cycle in humans, R. Soc. Open Sci. 3 (4) (2016) 160097.
[19] W. Bruine de Bruin, A.M. Parker, J. Strough, Age differences in reported social networks and well-being, Psychol. Aging (2019).
[20] T. David-Barrett, J. Kertesz, A. Rotkirch, A. Ghosh, K. Bhattacharya, D. Monsivais, K. Kaski, Communication with family and friends across the life course, PLoS One 11 (e0165687) (2016).