

A. Soboliev<sup>1</sup>, D. Lande<sup>2</sup><sup>1</sup> Institute of special communication and information security of National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine<sup>2</sup> Institute for information recording of National academy of science of Ukraine, Kyiv, Ukraine

## SOCIAL NETWORK NODES RANKING IN TERMS OF LOGARITHMIC FUNCTION OF ITS LINK WEIGHTS

Social networks are the basis of all interactions among its participants (usually people), which happen in the process of transmitting information. Lately this term is becoming more and more popular, but hardly anyone can really imagine how much it surrounds us. Social type networks are represented by means of graphs and node connections, which reflect real cooperation. It is necessary to conduct the detailed network analysis, evaluate the results by all specified standards and separate the most important nodes for ranking them in these networks. Existing ranking algorithms mainly evaluate everything in general, which does not allow to clearly recognise the consequence of nodes inter se. In the given article we provide the analysis of work educts of well-known node ranking algorithms (HITS, PageRank) and compare obtained data with expert network evaluation. Big amount of nodal connections in social networks and their various configuration in most cases do not allow to use the base type ranking algorithms, since the neglect of seemingly irrelevant connections induces false results. The base type algorithm HITS was adjusted for efficiency of the quasiheirarchic networks research. It allows to perform analysis and node ranking based on specified criteria (the amount of input and output connections inter se), which corresponds with the results of expert evaluation. It is displayed, that in some cases the received method offers corresponding with real social relations between subjects insights, and exponents of node authorships - with previously provided social roles. Received algorithm allows to evaluate and educe the most relevant nodes in social character networks. It can be used in various spheres, where social networks are formed.

**Keywords:** quasiheirarchic social character networks; network subjects; ranking algorithm; HITS method; PageRank method; node ranking; F-measure; expert evaluation.

### Introduction

Nowadays, when social networks are not just a composite of society, but have rather big influence on it, we should clearly understand all possible effluent risks and be ready for serious consequences in case of negligent attitude to this situation. Term “social networks” refers not only to online social networks, which are used as means of communication, but all connections, which happen among people (social subjects) for transferring information. It should also be noted that the network of social character is the network, nodes of which are social subjects, and connections are contacts, which are followed by information exchange among them. Defining the quasihierarchic network, one should consider the network close to hierarchic one with its hierarchy being violated with a small amount of additional connections.

**Work objective.** The objective of this work is the description and research of node ranking algorithm efficiency of quasihierarchic social networks, which is based on HITS algorithm modification as compared to other node ranking network algorithms.

**Analysis of recent researches and publications.** The research of social networks, definition of their organisation and value, analysis of the impact process on them are actively going on in the global IT sphere. Analysis of the mechanisms, which allow to control networks and conduct the necessary manipulations amidst informational intrusion of what is advisable for manipulator, starting with selling unnecessary goods and ending with cyberwars. Social network analysis process is very bothersome, as the wrong action can lead to an unexpected result or, in the worst case scenario, to the loss of control over this network.

### Exposition of basic research material

By the social character network ranking one should understand the nodes (social subjects) seriation process on the certain grounds, which allows to define the set nodes decisiveness inter se.

For visual display, this network is introduced by vectored graph, in which nodes are represented by subject tokens (social networks profile pages, phone numbers, electronic addresses), and connections between them are directed by links.

For mathematical representation such graph is displayed in the form of nodes adjacency matrix, and the value in this matrix box reflects the number of directed links.

According to HITS algorithm, two exponents are calculated for every network node:

- mediation index (hub);
- authorship index (auth);

The mediation index *hub* of  $A_i$  node of  $A_j$  network, which consists of  $n$  nodes equals nodes authorship vertical sum, which it refers to:

$$hub A_i = \sum_{j \leftarrow i} auth A_j,$$

Authorship index *auth* of  $A_i$  node of  $A_j$  network equals nodes mediations vertical sum, which it refers to:

$$auth A_i = \sum_{j \rightarrow i} hub A_j,$$

Presented formulas assign auxiliary iterated calculation algorithm of mediation and authorship indexes for all network nodes.

Talking about more complex networks (scale-free networks) with the number of nodes being over 1000 (big

organisations, etc.) the assumption about the social subjects, who serve as different ranks executives, is almost impossible without special operation algorithms [1].

Such characteristics as nodes degree distribution, which is defined as the probability of a node having a degree (node degree is the number of links connected to this node), resides in scale-free networks. It is a node degree exponential distribution network that is called scale-free, being most often observed in real life large networks [7].

It should be noted that the amount of connections in QSCN equals the amount of nodes. Given statement can be represented in the following mathematical view:

$$N_i = O(N_j),$$

where  $N_i$  – the amount of connections in the network;  
 $N_j$  – the amount of nodes in the network.

If examining QSCN one can observe clear (also plausible) one way link, it is undoubtedly which node in the pair has bigger rank.

In these networks 3 main kinds of relationships for a node pair are observed:

- leader - subordinate;
- equal rank nodes;
- the lower rank nodes have direct connection with the higher rank nodes.

When subject A directs subject B in social character network, it is offered to consider it as overall ratio weight, one of criteria of which is time index.

It is necessary to mention that a leader chaotically contacts a subordinate in case of problem statement necessity. Therefore a subordinate, with definite consistent periodicity, answers to a leader for outgoing order execution.

While estimating the nodes amount characteristics, the following fact takes place: a leader contacts a subordinate more often meaning that the significance of outgoing communication from a leader must exceed the incoming communication significance. [1]

It should be noted, that we examine only integrated and most widespread contact pattern between a leader and a subordinate. In real work this pattern can be completed with other specific characteristics, which will increase the efficiency of weight evaluation of node pair contacts.[6]

Sometimes it happens that a lower rank worker helps (during the work process) a worker of higher rank or a high rank executive gives an order to a worker of much lower hierarchy level. Such situations can lead to appearance of non-hierarchic connections in such type of network. These cases are rare, so we consider a network as quasihierarchic one. Meanwhile, using basic HITS algorithm for node ranking in the above-stated network can have little result, as even a weak and insignificant connection (disregarding its weight) can have a big influence on nodes portability and authorship indexes [4].

For node ranking in the social character networks we can use PageRank algorithms:

$$PR(A) = (1 + d) + d \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)},$$

where  $PR(A)$  – PageRank weight for page  $A$ ;  $d$  – damping index;  $PR(T_i)$  - PageRank weight for a page, which indicates page  $A$ ;  $C(T_i)$  - number of links from this page;  $n$  – pages, which refer to page  $A$ .

### Exposition of main results

For mathematical realisation of iterative node ranking conception it should be noted that the most influential subject in the network is the one having the biggest amount of outgoing connections with other network subjects.

For demonstration and evaluation of node ranking efficiency represented on Fig. 1, let us examine the QSCN consisting of 13 subjects with knowingly spaced certain connections, use the above-stated methods and compare the received results.

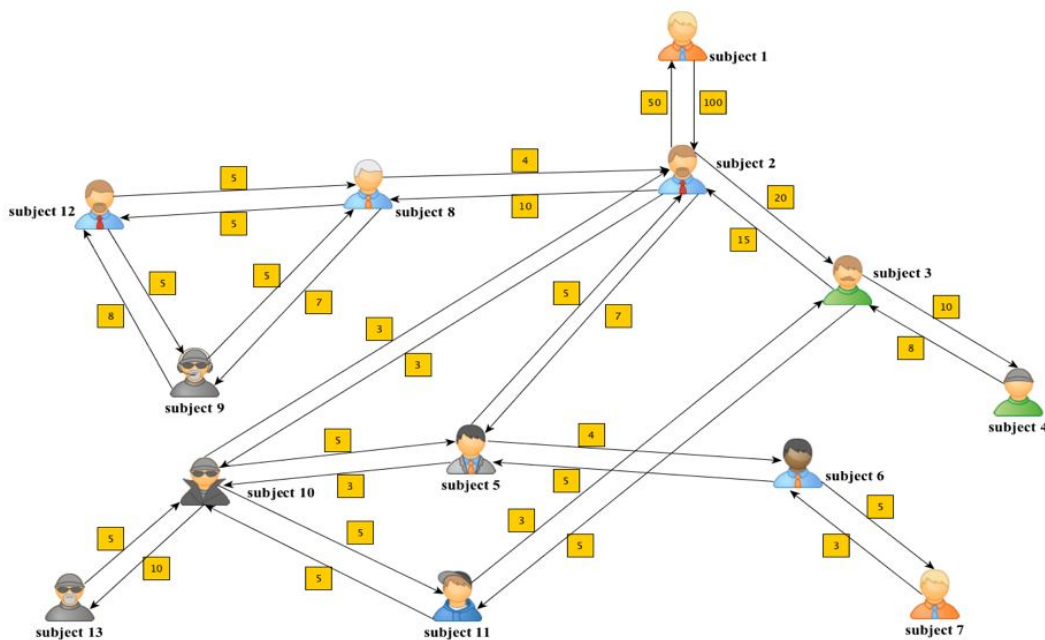


Fig. 1. Quasihierarchic social character network

When ranking the network with HITS algorithm, we ultimately receive two parameters: authorship and mediation indexes, which do not allow us to rank the network upon 2 criteria simultaneously. In order to solve this issue it is necessary to use metrics combining the information about authorship and mediation indexes of our algorithm. Let us use F-measure as such metrics:

$$F(A_i) = 2 / \left( \frac{1}{auth A_i} + \frac{1}{hub A_i} \right),$$

where  $F(A_i)$  – F-measure of  $A_j$  node of the network;  $hub A_i$  - mediation index;  $auth A_i$  – authorship index.

After analysing the received network ranking results according to Table 1, it is much in evidence, that authorship and mediation indexes are equal. It is due to feedback between nodes being observed in QSCN and basic HITS algorithm not providing the necessary results of ranking such networks. Also, comparing the work results of HITS and PageRank algorithms with given expert evaluation we can claim that the above-mentioned algorithms in their basic state are not effective for node ranking in QSCN.

PageRank index considers only incoming links and does not separate such concept as mediator. It severely influences the general result, as modest network connections can cardinaly change it.

It is offered to use HITS algorithm for method adaptation and consider edges weight when calculating every network nodes index.

To trace the efficiency of the above-mentioned statement, let us modify the HITS algorithm formulas using the multiplying by connections weight between nodes, resulting in:

$$hub A_i = \sum_{j \leftarrow i} auth A_j \cdot E_{ij},$$

where  $auth A_j$  – authorship index of node  $A_j$  of the network;  $hub A_i$  – mediation index of node  $A_j$ ;  $E_{ij}$  – connections weight between nodes  $A_i$  and  $A_j$ .

$$auth A_i = \sum_{j \rightarrow i} hub A_j \cdot E_{ij},$$

where  $hub A_j$  – mediation index of node  $A_j$  of the network;  $auth A_i$  – authorship index of node  $A_j$ ;  $E_{ij}$  – connections weight of between nodes  $A_j$  and  $A_i$ .

Table 1. The results of network node ranking with PageRank algorithms and basic HITS algorithm and expert evaluation

| Expert evaluation<br>(№ Node) | PageRank<br>(№ Node) | HITS          |              |            |
|-------------------------------|----------------------|---------------|--------------|------------|
|                               |                      | Auth (№ Node) | Hub (№ Node) | F (№ Node) |
| 2                             | 2                    | 2             | 2            | 2          |
| 1                             | 10                   | 10            | 10           | 10         |
| 3                             | 3                    | 5             | 5            | 5          |
| 8                             | 5                    | 8             | 8            | 8          |
| 5                             | 8                    | 3             | 3            | 3          |
| 10                            | 6                    | 11            | 11           | 11         |
| 11                            | 11                   | 1             | 1            | 1          |
| 9                             | 9                    | 9             | 9            | 9          |
| 12                            | 12                   | 12            | 12           | 12         |
| 4                             | 7                    | 6             | 6            | 6          |
| 13                            | 4                    | 13            | 13           | 13         |
| 6                             | 13                   | 4             | 4            | 4          |
| 7                             | 1                    | 7             | 7            | 7          |

Table 2. Results of network nodes ranking with modified HITS algorithm and experts evaluation

| Expert evaluation<br>(№ Node) | HITS          |              |            |
|-------------------------------|---------------|--------------|------------|
|                               | Auth (№ Node) | Hub (№ Node) | F (№ Node) |
| 2                             | 2             | 1            | 1          |
| 1                             | 1             | 2            | 2          |
| 3                             | 3             | 3            | 3          |
| 8                             | 8             | 5            | 8          |
| 5                             | 5             | 8            | 5          |
| 10                            | 10            | 10           | 10         |
| 11                            | 4             | 4            | 4          |
| 9                             | 11            | 9            | 11         |
| 12                            | 13            | 11           | 9          |
| 4                             | 12            | 12           | 12         |
| 13                            | 9             | 6            | 13         |
| 6                             | 6             | 13           | 6          |
| 7                             | 7             | 7            | 7          |

After the evaluation of the received work results of modified HITS algorithm according to Table 2 with experts evaluation data, we can claim that in general they are similar, but they do have irregularities. That is why the modified HITS method is ineffective during the general evaluation in QSCN. For detailed network evaluation it is necessary to analyse the received authorship and mediation results with expert evaluation data, as even insignificant connections can severely influence the ranking results and determining the most significant network nodes.

As weight definitions disposal irregularity can influence the results fidelity, their values must be decreased by multiplying them on some steadily increasing function, which is less steep that a linear one. In the context of the offered algorithm, we will use the modified HITS algorithm as such function. Then we will modify the formulas by multiplying them on hyperbolic logarithm of connections weight between nodes. The result is as follows:

$$hub A_i = \begin{cases} \sum_{j \leftarrow i} auth A_j \cdot \ln(E_{ij}), E_{ij} > 0; \\ 0, E_{ij} = 0, \end{cases}$$

where  $auth A_j$  – authorship index of node  $A_j$  of the network;  $hub A_i$  – mediation index of node  $A_i$ ;  $E_{ij}$  – connections weight between nodes  $A_i$  and  $A_j$ .

$$auth A_i = \begin{cases} \sum_{j \rightarrow i} hub A_j \cdot \ln(E_{ji}), E_{ji} > 0 \\ 0, E_{ji} = 0 \end{cases},$$

where  $hub A_j$  – mediation index of node  $A_j$  of the network;  $auth A_i$  – authorship index of node  $A_i$ ;  $E_{ji}$  – connections weight between nodes  $A_j$  and  $A_i$ .

For work efficiency evaluation of the received algorithm and the check concerning another parameter we will use degree of 1/3 of connections weight in the first case and hyperbolic logarithm in the other in modified  $auth A_j$  and  $hub A_j$  formulas and check their work results on QSCN.

During the research process of the received results set out in Table 3 of modified algorithm, we can observe considerable discrepancy in calculated values.

It is determined that, when using hyperbolic algorithm of connections weight in nodes, modified method provides the results, which correspond with the real social relations, while the authorship index showed the hierarchic subject dependence in the network.

Table 3. Results of network node ranking with modified HITS algorithms and experts evaluation

| Expert evaluation (№ Node) | HITS, 1/3     |              |            | HITS, ln      |              |            |
|----------------------------|---------------|--------------|------------|---------------|--------------|------------|
|                            | auth (№ Node) | hub (№ Node) | F (№ Node) | auth (№ Node) | hub (№ Node) | F (№ Node) |
| 2                          | 2             | 2            | 2          | 2             | 2            | 2          |
| 1                          | 1             | 1            | 1          | 1             | 1            | 1          |
| 3                          | 3             | 3            | 3          | 3             | 3            | 3          |
| 8                          | 5             | 10           | 5          | 5             | 10           | 8          |
| 5                          | 8             | 5            | 8          | 8             | 8            | 5          |
| 10                         | 10            | 8            | 10         | 10            | 5            | 10         |
| 11                         | 11            | 11           | 11         | 11            | 9            | 11         |
| 9                          | 12            | 9            | 9          | 12            | 11           | 9          |
| 12                         | 9             | 12           | 12         | 4             | 12           | 12         |
| 4                          | 13            | 4            | 4          | 13            | 4            | 4          |
| 13                         | 4             | 6            | 13         | 9             | 6            | 13         |
| 6                          | 6             | 13           | 6          | 6             | 13           | 6          |
| 7                          | 7             | 7            | 7          | 7             | 7            | 7          |

It should be noted that given modified algorithm with the value of connections weight hyperbolic logarithm was used for the research of 20 different quasihierarchic networks, and the received values results represented set facts.

### Conclusions

The notion of the quasihierarchic social character networks is introduced and connections character in such networks is researched in this work.

The efficiency research of basic and modified HITS algorithms for node ranking of quasihierarchic

social character networks was conducted in comparison to PageRank algorithm in the given network.

When used practically, it is shown that modified algorithm provides results corresponding with real social relations, and nodes authorship indexes - their hierarchic importance inter se.

The received algorithm was used for research of 20 different social networks and, in the process of work, showed pretty exact results, which represented set facts.

Modified HITS algorithm is effective for social network research in the process of cyberwar and for network node ranking in general.

## REFERENCES

1. Lande, D.V and Nechaev, A.O. (2015), "Algorithm for ranking of nodes of quasi-hierarchical networks of social character", *Problems of informatization and management: a collection of scientific works*, No. 1 (49), NAU, Kyiv, pp. 46-50.
2. Lande, D.V., Snarsky, AA and Bezsudnov, I.V. (2009), *Internet: Navigation in complex networks: models and algorithms*, Librokom (Editorial URSS), Moscow, 264 p.
3. Sulema, O.K. and Lande, D.V. (2015), "Finding the optimal hierarchy in the quasi-hierarchical graph according to the criteria of centrality", *Registration, storage and processing of data*, Vol. 17, No. 4, pp. 3-10.
4. Bargh, J.A., Chen, M. and Burrows, L. (1996), "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action", *Journal of Personality and Social Psychology*, Vol. 71, No. 2, pp. 230-244.
5. Laura, L.S. and Gianluigi, Me. (2017), "Searching the Web for illegal content: the anatomy of a semantic search engine", *Soft Computing*, pp. 1245-1252.
6. Liu, Y.Y., Jean-Jacques Slotine, J.J. and Barabasi, A.L. (2012), "Control centrality and hierarchical structure in complex networks", *PLOS ONE*, Vol. 7, No. 9, pp. 1-7.
7. Cohen, David (2002), "All the World's a Net", *Newscientist*, Vol. 174, No. 2338, April 13, 2002, pp. 24-29.
8. Amy, N. Langville and Carl, D. Meyer, (2013), "Google's PageRank and Beyond: The Science of Search Engine Rankings", *PRINCETON UNIVERSITY PRESS*, pp. 115-121.
9. Scott, John (2017), "Social network analysis", CPI Group(UK) LTd, Croydon, 4<sup>th</sup> edition, pp. 57-89.
10. Malcolm, R. Parks, (2016), "Personal relationship and personal network", *Malcolm R. Parks*, New York, 1<sup>st</sup> edition, pp. 2-14.
11. Stephen, P. Borgatti, Martin, G. Everett and Jeffrev, C Johnson (2018), "Analyzing Social Networks", *C&M Digitals (P) Ltd, Crennai, India, Printed UK*, pp. 13-25.

Надійшла (received) 24.04.2018

Прийнята до друку (accepted for publication) 11.07.2018

### **Ранжування мережі соціального характеру із врахуванням логарифмічної функції її зв'язків**

А. М. Соболев, Д. В. Ланде

Соціальні мережі є основою всіх взаємодій між учасниками (як правило, людьми), які відбуваються в процесі передачі інформації. Останнім часом цей термін стає все більш популярним, але навряд чи хтось може уявити, наскільки сильно він оточує нас. Мережі соціального характеру зображають за допомогою графів та зв'язків між вузлами, що відображають реальну взаємодію. Для ранжування вузлів в даних мережах необхідно провести детальний аналіз мережі, оцінити отримані результати за всіма заданими критеріями та виділити найбільш впливові вузли. Існуючі алгоритми ранжування переважно оцінюють все в цілому, що не дозволяє чітко визначити впливовість вузлів між собою. У даній статті приводиться аналіз отриманих результатів роботи відомих алгоритмів ранжування вузлів (HITS, PageRank) та порівнюються отримані дані з експертною оцінкою мережі. Велика кількість зв'язків між вузлами та їх різноманітна конфігурація в більшості випадків не дозволяє використовувати алгоритми ранжування базового типу, оскільки нехтуючи начебто несуттєвими зв'язками, виникають хибні результати. Базовий алгоритм HITS був модифікований для ефективного дослідження квазієрархічних мереж. Показано, що отриманий алгоритм надає у ряді випадків результати, які відповідають реальним соціальним відношенням між суб'єктами, а показники авторства вузлів – попередньо наданим соціальним ролям. Отриманий алгоритм дозволяє оцінювати та виявляти найбільш релевантні вузли в мережах соціального характеру. Він може бути використаний у різних сферах, де формуються соціальні мережі.

**Ключові слова:** квазієрархічні мережі соціального характеру; суб'єкти мережі; алгоритм ранжування; алгоритм HITS; алгоритм PageRank; рейтинг вузла; F-міра; експертна оцінка.

### **Ранжирование сети социального характера с учетом логарифмической функции ее связей**

А. М. Соболев, Д. В. Ланде

Социальные сети являются основой всех взаимодействий между участниками (как правило, людьми), которые происходят в процессе передачи информации. В последнее время это определение становится все более популярным, но вряд ли кто-то может представить, насколько сильно он окружает нас. Сети социального характера изображают с помощью графов и связей между узлами, отражающими реальное взаимодействие. Для ранжирования узлов в данных сетях необходимо провести детальный анализ сети, оценить полученные результаты по всем заданным критериям и выделить наиболее влиятельные узлы. Существующие алгоритмы ранжирования преимущественно оценивают все в целом, что не позволяет четко определить влиятельность узлов между собой. В данной статье приводится анализ полученных результатов работы известных алгоритмов ранжирования узлов (HITS, PageRank) и сравниваются полученные данные экспертной оценке сети. Большое количество связей между узлами и их разнообразная конфигурация в большинстве случаев не позволяет использовать алгоритмы ранжирования базового типа, поскольку пренебрегая якобы незначительными связями, возникают ложные результаты. Базовый алгоритм HITS был модифицирован для эффективного исследования квазиерархических сетей. Показано, что полученный алгоритм предоставляет в ряде случаев результаты, которые соответствуют реальным социальным отношениям между субъектами, а показатели авторства узлов - предварительно предоставленным социальным ролям. Полученный алгоритм позволяет оценивать и выявлять наиболее релевантные узлы в сетях социального характера. Он может быть использован в различных сферах, где формируются социальные сети.

**Ключевые слова:** квазиерархические сети социального характера; субъекты сети; алгоритм ранжирования; алгоритм HITS; алгоритм PageRank; рейтинг узла; F-мера; экспертная оценка.