Converting network–unlike data into complex networks: problems and prospective

A A Tikhomirov¹, O G Berestneva², E Mokina², N Kinash³, M Kuklina³, A I Trufanov³, A Rossodivita⁴, V Kuklina⁵, I Bilichenko⁶, V Bogdanov⁷

¹ CIS Collaboration Center, Inha University, 100 Inharo, Nam-gu, Incheon, 22212, Republic of Korea
² Division for Information Technology, Tomsk Polytechnic University, 30, Lenina ave., Tomsk, 634050, Russian Federation
³ Institute of Information Technology and Data Science, Irkutsk National Research Technical University, 83, Lermontova st., Irkutsk, 664074, Russian Federation
⁴ Medical and University Center Luigi Sacco Hospital, 74, Via GB Grassi, Milan, 20157, Italy
⁵ Institute for European, Russian, and Eurasian Studies, George Washington University, 1957, E st. NW, Washington, D.C., 20052, United States of America
⁶ Laboratory of Physical Geography and Bio-geography, V.B. Sochava Institute of Geography, 1, Ulan-Batorskaya st., Irkutsk, 664033, Russian Federation
⁷ Laboratory of Cartography, Geoinformatics and Remote Sensing, Methods, V.B. Sochava Institute of Geography, 1, Ulan-Batorskaya st., Irkutsk, 664033, Russian Federation
⁸ ogb6@yandex.ru

Abstract. Often network science with complex networks as its basic entity has attracted scientific societies with their diverse practical capacities. Some entities (objects, processes, and data) having their built-in web nature might be considered as networks seamlessly. Contrary, network mapping for Network –Unlike Data (NUD), i.e. images and time series, is extremely complicated and manifold, so that explorers face with a tough problem which converting algorithm they should apply. We put in central focus separating data properties in line with their scale diversity – in distance, time, and nature and suggested a three step algorithm (scale-based one) to map NUD into complex networks. The algorithm was applied to networkalize two types of geographic maps of Olkhon district, Baikal Natural Territory, Irkutsk region, Russian Federation. It was underlined that the algorithm models coarse-graining and area-like linking and forms thoroughly output structures of really complex topologies with intrinsic scale-free and small world properties. In addition to simple examples transformation of NUD into multiplex networks is considered as a promising approach to study more complex systems of the real world. Networkalization concerned challenges in extracting the pertinent information from huge data resources conveyed by a network imprint for each file is also discussed.
1. Introduction

Network science with complex networks as its basic paradigm has become in demand by scientific societies through their diverse theoretical and practical capacities [1]. Scholarly have modelled successfully synthetic and real Data in formats of network structures and elaborated powerful instruments for further effective analysis with support by Graph Theory, Probability Theory, Linear Algebra, and Fractal Geometry.

It is of value that the more number of actors in a system are comprised, the more one argues utilization of a network approach to study the system. Often some systems with their intrinsic natures which are similar to web ones have been interpreted as networks, e.g. facilities and channels for transportation, communication, water supply and other assets of critical infrastructure. Many activities connected with travels such as tourism [2], commuting [3], trading [4] also might be modelled by networks seamlessly. Social relations [5-6] and concomitant spread processes for information and infection diseases [7] compose entities of a network profile. It is of sense to call such systems, processes, and data Network-Like ones [8]. The exploration of network-like kingdom [9] has seized scholar interests within diverse fields, theoretical and practical both [10-11]. Even more than two decades passed since network science birth, still experts inheriting graph theory formalism shuffle the words of these two interdependent domains.

At the same time it is of value to draw a boundary between mathematics and network science just to clarify which domain the issues focus on. Thus if one uses the terms “connectivity”, “nodes”, and “links”, it elucidates that network science is represented, not mathematics within theory of graphs. Network science defends the boundary with graph theory by taking into account metrics, models, and ontologies based on observation of real world.

An instrument to extract Network-Like Data from images automatically has been proposed in [9]. However, prior to mapping even Network-Like Data, such as concomitant to transportation systems should be thoroughly analyzed. Scholars must take into account the fact that not all complex network concepts are applicable to the transportation context and that some variations are acceptable [12].

Thus several models are usually considered for transportation systems, at least those of L- and P- [13] as manifested in [14] for Russian railways (Fig. 1).

![Figure 1](image-url)

**Figure 1.** Russian railways mapped as network models in L-space (a) and P-space (b) respectively.
Concerning Network –Unlike Data (NUD) of Spatial and Temporal character, their network interpretations are extremely manifold so that one encounter with a tough problem: which converting algorithm it is of sense to apply for significant result.

2. Related works
In general the technique of mapping Temporal Data -a time series into complex networks is represented by 3 next steps: the segmentation according to the specified time length, coarsening the ranges, linking the ranges with appropriate values for link weights.

One can find variety of ways to convert time series into complex networks, among them one may list:

- Coarse-graining algorithm based on statistics of segments [15];
- Visibility Graph (VG), including Natural Visibility Graph and Horizontal Visibility Graph [16], and also Multiplex Visibility Graph [17];
- Recurrence networks RN [18]
- Local Sort algorithm [19].

At the same time there are few dependable practices that utilize real complex networks for analysis and classification of Spatial Data (i.e. maps or images).

In the work [20] it was demonstrated that when a plane is randomly divided into intersecting adjacent fragments it is possible to build a scale-free network in which the nodes are fragments, and connections conform common boundaries between fragments. This forms a so-called “boundary” approach.

Second approach might be called as “pixel vicinity” algorithm. Just to treat digital image as a complex network: each pixel represents a node; those are linked or not according to their vicinity in location and similarity in intensity. In addition the set of complex network metrics should be studied to model images adequately.

Third approach is so called “image visibility graphs (IVG)” method with a row of image visibility graphs (IVG/IHVG): those have been introduced as simple algorithms by which scalar fields can be mapped into graphs [21].

Prior to utilization of any approach to map NUD into a network one should aware that the principal objective is to preserve key properties encoded in the data to the most extent. Comparison of the approaches might be qualitative or quantitative (e.g. based on entropy metric).

3. Model
Further it seems reasonable to call any transformation of NUD into a network with very simple and clear term: “networkalization”. Contrary to well-known algorithms of mapping NUD (presented as Spatial and Temporal Data) into complex networks based on the only scale, this study focuses on separation of the data properties in line with their scale diversity – in distance, time, and nature.

For Spatial Data reflected in a map (e.g. geographic map) we propose a three scale-based technique (partially similar to our previous scope [22]), the one with specific coarsening scheme applied to selected areas.

First, for all chosen areas regular (or irregular) networks are inserted so that each network size depends on area square S and its form (Fig.2).

![Figure 2. Step#1 of the algorithm that maps Spatial Data grains into complex networks: triangulation.](image-url)
Number of nodes within the area $N_{\text{local}} \sim S$

Two nodes are connected if distance between them is not greater than a threshold ($< d_{\text{cut}}$). Alternate linkage might be lattice grid. Also it makes sense to build the local network as a regular grid in case of regular node seeding.

At the second step, areas as coarsened entities are aggregated by one or several inside nodes (so-called central nodes), numbers of those depends on the corresponding local (inner) network size. It is reasonable to consider this super network of central nodes as dense or even complete one. Also the central nodes of different areas might be linked in case of common border between two areas (Fig.3).

Number of such border links depends of the length of the common border: Number of area grains (central nodes) f: $\text{Scale} \rightarrow N_{\text{grain}}$.

Number of links between central nodes in grains i and j respectively with common border length $L(i,j)$ is:

$$M_{\text{grain } i,j} \sim \frac{L(i,j)}{\text{Sum} (L(i,j))} ,$$

where $L$ is a common border length for i and j grains.

**Figure 3.** Step#2 of the algorithm that maps Spatial Data into a complex network: links for coarsened areas.

Third, grains with similar nature have a sufficient probability to be connected even with no common border and in case when their Euclidian distance is significant (Fig.4). Let link probability between similar nature grains $P(\text{Link } S_{\text{grain } i,j})$ be equal 1, if Manhattan distance(i,j) is not greater than a corresponding threshold value $b_{\text{cut}}$.

**Figure 4.** Step#3 of the algorithm that maps Spatial Data into a complex network: a link for areas of similar nature.

4. Findings

Aforementioned scale-based approach was utilized to networkalize landscape and land use maps representing Olkhon district, Baikal Natural Territory (BNT), Irkutsk region, Russian Federation. The results are portrayed on Fig.5 and Fig.6 respectively.
The study showed that coarse-graining and area-like connecting thoroughly form scale-free and small world properties, which control reducing the average shortest path length [1] in output networks, thus imbedding into their topologies real complexity. It is of value that generalization of the maps excluded some details and changed network imprints but transform network topologies insignificantly.

As it was shown semi-qualitative and semi-quantitative networkalization approach utilized in the study and consequent scale-based technique provide preserving key information underlying in NUD while mapping the latter into a network.

5. Discussion

Being in line with Gartner’s IT Glossary and similar to its terms of Digitization[23] and Digitalization [24], we refer to “networkalization” as “the use of network scope to add a scientific or a business network model and thus provide new advances and revenue and knowledge- or value-
producing opportunities; it is the process of moving to a network domain and network society in whole”. Contrary, “Networkization” is “the process of changing from natural structural form directly to network”. In other words networkization considers a natural entity (system, process and object) just to change concomitant Data into a network form without any structural changes to the entity itself. Thus an expert can “networkize” only network-like data, but for Network-Unlike Data he/she must utilize networkalization.

The examples of networkalization presented above are rather simple.

Transformations of NUD into multilayer (multiplex) networks [25] or into stem networks [26] might be considered as promising approaches to study more complex systems of the real world thoroughly.

A multiplex per se and a stem network are compact models that stratify interactions among actors of the same type (both for netwokization and networkalization cases). These models have been molded as prominent platform for exploring complex systems across variety of domains covering living organisms, technological and natural objects, human societies and others [27].

Stored and being registered Big Data in the format of texts (literature, scientific works, speeches) and spatio-temporal dynamics (audio and video records, represented in medicine, ecology, geology, meteorology and other domains) requires effective and efficient instruments to study and utilize that for national and international good. Networkalization implies impressive challenges in extracting the pertinent information from huge data resources conveyed by a network (or multiplex network) imprint for each file. Such a kind of augmented indexing Big Data with network metrics provides high performance searching within given domain. It seems prospective for detecting signs of unwanted regional and global processes caused by man-made and natural factors.

6. Conclusions
To conclude with it is necessary to note that the performed research reveals that it is of value to take into account the nature and specificities of Network-Unlike Data while considering them as networks. This context arms a researcher with a scope that gives a plus while dealing with real systems not to lose underlying information and conserve that for analysis and utilization in the domain of practice. If networkalize NUD properly the information hidden in data might be thoroughly inherited by complex network properties reflected in a set of numerous metrics to be revealed further through scale–based processing.

We suppose the scale-based approach of mapping Spatial and Temporal Data into complex networks might be valuable for theory and practice both.

Acknowledgments
The study was carried out with the partial financial support of the Russian Foundation for Basic Research (RFBR) within scientific project No. 18-07-00543.

References
[1] Vespiagnani A 2018 Twenty years of network science Nature 558 528–529 doi: 10.1038/d41586-018-05444-y
[2] Baggio R 2013 Studying complex tourism systems a novel approach based on networks derived from a time series Proc XIV April Int. Academic Conf. on Economic and Social Development (Moscow) (Preprint arXiv:1302.5909 [physics.soc-ph])
[3] De Montis A, Chessa A, Campagna M, Caschili S, and Deplano G 2010 Modeling commuting systems in a complex network analysis: A study of the Italian islands of Sardinia and Sicily Journal of Transport and Land Use 2 3(4) 39-55 doi: 10.5198/jtlu.v2i3.14
[4] Lee J W, Maeng S E, Ha G-G, Lee M H, and Cho E S 2013 Applications of complex networks on analysis of World Trade Network Journal of Physics: Conference Series 410 012063 doi: 10.1088/1742-6596/410/1/012063
[5] A. Semenov A, Mantzaris A V, Nikolaev A, Veremyev A, Veijalainen J, Pasiliao E L and Boginski V 2018 Exploring Social Media Network Landscape of Post-Soviet Space IEEE Access 7 411-426 doi: 10.1109/ACCESS.2018.2885479
[6] Gadek G, Pauchet A, Brunessaux S, Khelif K and Grilheres B 2018 AI techniques to analyse a social network on text, user and group level : application on Galaxy2. 4ème conférence sur les Applications Pratiques de l’Intelligence Artificielle APIA2018 (Nancy, France) (Preprint hal-01830922)

[7] Wu Q and Zhu W 2017 Toward a generalized theory of epidemic awareness in social networks Journal of Physics: Conference Series 280(05) 1

[8] Newman M 2003 The Structure and Function of Complex Networks SIAM Review 45 167–256

[9] Dimberger M, Kehl T and Neumann A 2015 NEFI: Network Extraction From Images Sci Rep 5 15669

[10] Zaidi F 2010 Analysis, Structure and Organization of Complex Networks. Networking and Internet Architecture [cs.NI] (Université Sciences et Technologies - Bordeaux I) (Preprint hal-tel-00542703)

[11] Liu X F and Tse C K 2015 A General Framework for Complex Network Applications Preprint arXiv:1507.05687 [physics.soc-ph]

[12] Zanin M, Sun X and Wandelt S 2018 Studying the topology of transportation systems through complex networks: handle with care Journal of Advanced Transportation 2018(1) 1-17 doi:10.1155/2018/3156137

[13] Derrible S, Kennedy C 2009 Transportation Research Record Journal of the Transportation Research Board 21(12) 17–25

[14] Tikhomirov A, Rossodivita A, Kinash N, Trufanov A, Berestneva O 2017 General topologic environment of the Russian railway network Journal of Physics: Conference Series 803(1) 012165 doi:10.1088/1742-6596/803/1/012165

[15] Zhang Z, Xu J and Zhou X 2019 Mapping time series into complex networks based on equal probability division AIP Advances 9 015017 doi:10.1063/1.5062590

[16] Lacasa L, Luque B, Luque J and Nuno J C 2009 The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion Preprint arXiv:0901.0888 [physics.data-an] doi:10.1209/0295-5075/86/30001

[17] Lacasa L, Nicosia V and Latora V 2015 Network structure of multivariate time series Scientific Reports 5(1) 15508 doi:10.1038/srep15508

[18] Marwan N, Donges J F, Zou Y, Donner R V and Kurths J 2009 Complex Network Approach for Recurrence Analysis of Time Series Physics Letters A 373 4246-54 doi:10.1016/j.physleta.2009.09.042

[19] Sun X, Small M, Zhao Y and Xue X 2014 Characterizing system dynamics with a weighted and directed network constructed from time series data Chaos: An Interdisciplinary Journal of Nonlinear Science 24 024402 doi:10.1063/1.4868261

[20] Hassan M K, Hassan M Z and Pavel N I 2010 Scale-free network topology and multifractality in a weighted planar stochastic lattice New Journal of Physics 12 093045

[21] Iacovacci J and Lacasa L 2018 Visibility graphs for image processing Preprint arXiv:1804.07125 [physics.data-an]

[22] Trufanov A, Kinash N, Tikhomirov A, Berestneva O, Rossodivita A 2017 Image Converting into Complex Networks: Scale-Level Segmentation Approach. Proc. IV Int. Conf. “Information technologies in Science, Management, Social sphere and Medicine” (ITSMSSM 2017), Series: Advances in Computer Science Research (ACSR) (Atlantis Press) 72 417-422

[23] Digitization Gartner Glossary [online at https://www.gartner.com/en/information-technology/glossary/digitization , accessed 9 Sept 2019]

[24] Digitalization Gartner Glossary [online at https://www.gartner.com/en/information-technology/glossary/digitalization , accessed 9 Sept 2019]

[25] Kivela M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 Multilayer networks Journal of Complex Networks 2(3) 203–271 (Preprint arXiv:1309.7233 [physics.soc-ph])

[26] Urazova N, Kuklina M, Kotelnikov N, Kaymonova O, Trufanov A 2019 Application of the Theory of Complex Networks in Analysis of Tourist Infrastructure Proc. of the Int. Conf. on
Humanities and Social Sciences: Novations, Problems, Prospects (HSSNPP 2019) (Atlantis Press) 720-724 doi: 10.2991/hssnpp-19.2019.137

[27] Hanteer O, Interdonato R, Magnani M, Tagarelli A and Rossi L 2019 Community Detection in Multiplex Networks Preprint arXiv:1910.07646 [cs.SI]