

Revealing effective classifiers through network comparison

LAZAROS K. GALLOS and NINA H. FEFFERMAN

Department of Ecology, Evolution, & Natural Resources, Rutgers University - New Brunswick, NJ 08901, USA, and DIMACS, Rutgers University - Piscataway, NJ 08854, USA

received 10 September 2014; accepted in final form 8 October 2014
published online 24 October 2014

PACS 89.75.Fb – Structures and organization in complex systems

PACS 89.75.Da – Systems obeying scaling laws

PACS 87.23.Ge – Dynamics of social systems

Abstract – The ability to compare complex systems can provide new insight into the fundamental nature of the processes captured, in ways that are otherwise inaccessible to observation. Here, we introduce the n -tangle method to directly compare two networks for structural similarity, based on the distribution of edge density in network subgraphs. We demonstrate that this method can efficiently introduce comparative analysis into network science and opens the road for many new applications. For example, we show how the construction of a “phylogenetic tree” across animal taxa according to their social structure can reveal commonalities in the behavioral ecology of the populations, or how students create similar networks according to the University size. Our method can be expanded to study many additional properties, such as network classification, changes during time evolution, convergence of growth models, and detection of structural changes during damage.



Copyright © EPLA, 2014

Advances in quantitative methods for network analysis have found many applications in a startling diversity of fields [1]. As in the progression of many quantitative tools, while initial efforts to use network analysis were mainly descriptive [2], research then advanced to focus on using them as predictive tools, isolating particular characteristics that can provide insight into the system of interest. However, the richest and most interesting level of investigation from new metrics frequently arises when they are ultimately used to make comparisons across systems, discovering which characteristics are shared and which are not. The ability to compare systems has always been a strong driving force in science [3].

Currently, there is not a rigorous definition of network similarity. This allows similarity to be as broadly interpreted as just one single quantity averaged over the entire system —*e.g.* networks with the same average degree— or it can be extremely restrictive, *e.g.* node-to-node correspondence in identical networks. Obviously, no one property can fully characterize a network: for instance, networks can be structurally very different even if they have the same degree distribution but different clustering coefficient, or different modularity, etc. It is not known how many and which properties should be combined to construct a weighted index of similarity. Therefore, current research has been directed to alternative methods.

For example, correlation analysis has been used to detect similarities in financial and biological systems [4–7]. Structure-based methods, such as motif comparison [8] or graphlet comparison [9], are based on the idea that if we continuously isolate parts of the network and find the same patterns to occur in the same frequency in two networks, these networks will have a higher probability of being “similar” to each other. However, there are many practical constraints that render these techniques incapable of handling larger networks or larger motifs [10]. The most recent advance in the field [11] introduced a novel concept in which the network is broken down in communities at different scales and the comparison is based on network modularity properties. The question of similarity under this method becomes: “How similar is the modular structure of the networks?”

Here we introduce a measure to detect similarity based on direct topological properties: Topological Analysis of Network subGraph Link/Edge (tangle) Density. Many of these properties can be captured by the distance from a tree structure at different length scales—which means that all tree structures will be deemed equivalent even if they are different structures, *e.g.* a scale-free tree *vs.* an ER tree. The method combines the insight of motifs, simplified for efficiency, and focuses on microscopic structure compared to the mesoscopic approach of modularity

comparison in Onnela *et al.* [11]. Where the advantage of the motif method is that it takes into account the local configuration of the links, if we relax the motif requirement for exactly matching patterns we can use the links density as our metric. The basic foundation of our method is to calculate how the density of links behaves at different scales across the network. It is also possible to use other properties instead of density, such as the local degree distribution or clustering coefficient, but the crucial step is the sampling of the connected subgraphs. For example, the method can be easily extended to cover weighted networks, by substituting the number of links in the subgraph with the total sum of weights in the same subgraph. The general interpretation of the method is that if two networks are found similar then, given a part from one network, we cannot distinguish from which network the part was extracted.

The crux of this method is to capture how many affiliations we expect to find when we isolate any given size of connected sub-group. The concept is the following: Consider a connected group of 10 students, which is randomly selected from a class of 100 students. If you are in this group, how many direct friends do you expect to find in this sample, or in other words what is the average edge density in the group? We define this to be the 10-tangle density (or n -tangle, for any n). If we construct the histogram of densities from different samples over different sizes, then we can compare these distributions in two different networks, and we can know the extent of association in a group of a given size independently of the pattern formed in each subgroup. In this way, our method bypasses the need to determine direct node-to-node correspondence [12], while still capturing node-level properties of the network for comparison.

Formally, we define the n -tangle method in the following way. In a graph $G(V, E)$ comprising a set V of nodes and a set E of edges we isolate all possible connected induced sub-graphs $G^{\text{in}}(V_n, E_n)$. The condition for these sub-graphs is that they should include exactly n nodes ($|V_n| = n$) and the subset E_n of E should include all e_n links among those n nodes in G . For each subgraph we define the n -tangle density, t_n , as the normalized edge density of this subgraph, *i.e.* the fraction of existing over all possible links, after we remove the $n - 1$ links that are needed for connectivity:

$$t_n = \frac{e_n - (n - 1)}{n(n - 1)/2 - (n - 1)} = \frac{2(e_n - n + 1)}{n^2 - 3n + 2}. \quad (1)$$

It is important that the size of the n -tangle remains much smaller than the network size N , $n \ll N$, so that the sampled subgraphs are statistically independent of each other. To include the considerably inhomogeneous character of the local structure in networks, we consider the n -tangle distribution $P(t_n)$ of all G^{in} (figs. 1(A), (B)). This distribution represents the “signature” of a network at a given subgraph size n . We repeat this process for

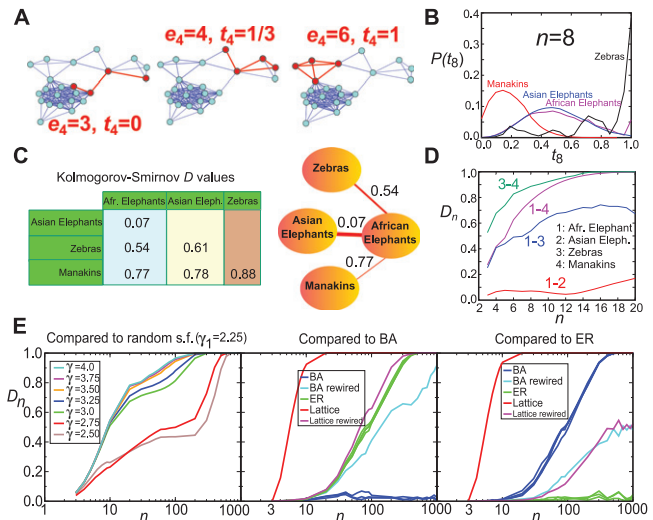


Fig. 1: (Color online) The n -tangle method. (A) We randomly sample connected induced subgraphs of n nodes and calculate their normalized link density t_n . (B) We construct the n -tangle histogram $P(t_n)$ for a given value of n (the example shows the 8-tangle distribution for 4 animal social networks). (C) We calculate the distance between any two distributions I and J through, *e.g.*, a Kolmogorov-Smirnov statistic $D_n(I - J)$. These D values are used as the distance between the original networks and can be mapped to a minimum spanning tree (shown here for the four networks), a hierarchical tree, or a threshold-based network. (D) Variation of the distance between these 4 networks as a function of the subgraph size n . (E) The distance of a random scale-free network with a degree exponent γ from a network with $\gamma_1 = 2.25$ increases monotonically as we increase the value of γ (left). Similarly, the n -tangle distance among random Barabási-Albert networks (center) or random Erdős-Rényi networks (right) is close to zero, while distances with other model networks are significantly higher.

all different subgraph sizes n , resulting to potentially different signatures as we vary n . We can then compare the degree of similarity of two networks A and B at a given scale by a simple Kolmogorov-Smirnov statistic [13], $D_n(A - B) = \sup |F_A(t_n) - F_B(t_n)|$, where $F_A(t_n)$ is the corresponding n -tangle cumulative probability in network A and \sup denotes the supremum value (fig. 1(C)). Since the full comparison involves all subgraph sizes, this method can reveal how two networks can be similar at a local scale, while at a larger scale they may exhibit different structures, allowing both global network comparison and local analysis of the scale at which similarity may be greatest (fig. 1(D)).

Our approach avoids the inherent constraints of motif [8] or graphlet [9] based methods [14], by ignoring the costly calculation of the specific pattern created by the group and instead placing emphasis on the density of the group, *i.e.* a single number. Therefore, the exponential increase in the number of patterns as a function of group size, which limits those techniques to very small-size patterns, does not influence the applicability of our method to larger sub-graphs. Using this method, we only need to keep the

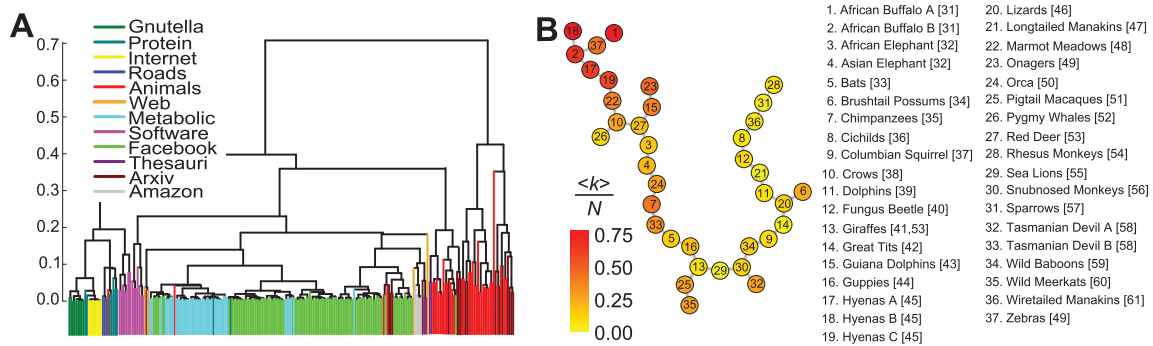


Fig. 2: (Color online) (A) Hierarchical tree of 236 networks from different fields, based on the n -tangle distance (here $n = 5$). We used the UPGMA (Unweighted Pair-Group Method using Arithmetic Averages) hierarchical clustering method [15]. Colors represent networks in the same family, as indicated in the index. (B) The Minimum Spanning Tree of animal networks, based on the n -tangle distance ($n = 12$). The species color corresponds to varying levels of normalized degree $\langle k \rangle / N$ and separates nicely the species. The citations in the index indicate the source of the corresponding datasets.

number of links for each configuration, which makes the calculation and storage very fast. Even though the computational complexity of the n -tangle method does increase with the subgraph size, the connectedness of bigger social groups can be probed at practically any size n , through a fast sampling method. We used a simple Monte Carlo method to sample a large number of configurations, where we repeatedly selected random subgraphs and calculated the links within.

Our method can successfully detect changes in standard network properties. In fig. 1(E) we compare a series of random scale-free networks created by the configuration model with a similar network with degree exponent $\gamma_1 = 2.25$. The networks become more distant as the exponent of these networks increases, demonstrating that the method can separate similar structures with different parameters. Similarly, we compare a number of networks to a sample Barabási-Albert (BA) network. Different realizations of BA networks are found to be at almost zero distance from each other, but a randomly rewired BA network has a different structure. Similarly, lattices and Erdős-Rényi (ER) networks are also far from the BA network. Analogous results are found when we compare these model networks with an ER network.

We demonstrate the n -tangle method first by comparing 236 network structures of different origins (described in the appendix). The hierarchical tree in fig. 2(A) indicates that networks from the same family tend to cluster with each other. For example, friendship networks in facebook are, in general, closer to each other than to, *e.g.*, animal social networks, which also tend to be detected as similar. We consider this natural separation as a simple verification test for the method.

A more interesting problem is to detect network similarities in systems from within the same family. For example, we can construct a “phylogeny” of animal species based on their social structure [16]. In this way, we explore whether species with similar descriptive characterizations in behavioral ecology do, in fact, exhibit similar social

structures [17]. We analyze empirically determined contact affiliation networks of 33 animal species. In molecular biology, phylogenetic trees can be constructed from evolutionary distance (pairwise distances between sequences). Here, rather than using species genetic data, our input data are the pair-wise distances of the n -tangle method. We are therefore able, using our analysis, to determine whether or not a meaningful cluster results from a choice of a particular facet of the system. In this example, we find that the normalized average degree, *i.e.* $\langle k \rangle / N$, is able to generate clear clustering by n -tangle analysis. This result of our method can provide the first insights into whether qualitatively similar social classifications in fact yield similar population-level networks of interaction across species (for example, do all dominance hierarchies yield similar social structures for the entire population?). This is a critical next step in understanding animal social systems.

The n -tangle method can also be used to isolate key network features that enable classification of networks. In fig. 3 we present the n -tangle connectivity trees resulting from a) Facebook friendship networks in 100 universities in 2005 [18], b) arxiv.org co-authorship [19] in 17 different fields, and c) software code in 14 different projects [20,21]. For the Facebook friendship, there is no clear clustering with the average degree, but when we consider student enrollment, then we discover a similarity between networks at universities of similar size, at all sizes. The n -tangle method therefore enables us to obtain meaningful sociological insight into the process, where students create online friendships according to the size of the pool of possible connections, even though the average number of friendships is much smaller than the pool size. It may therefore be that the fundamental nature of the social activities and experience is shaped by the total size of the university, even though that number can be significantly larger than the size of the average friend-group. For the case of co-authorship, on the other hand, the classification of networks according to the network size does not work well. We instead discover that the

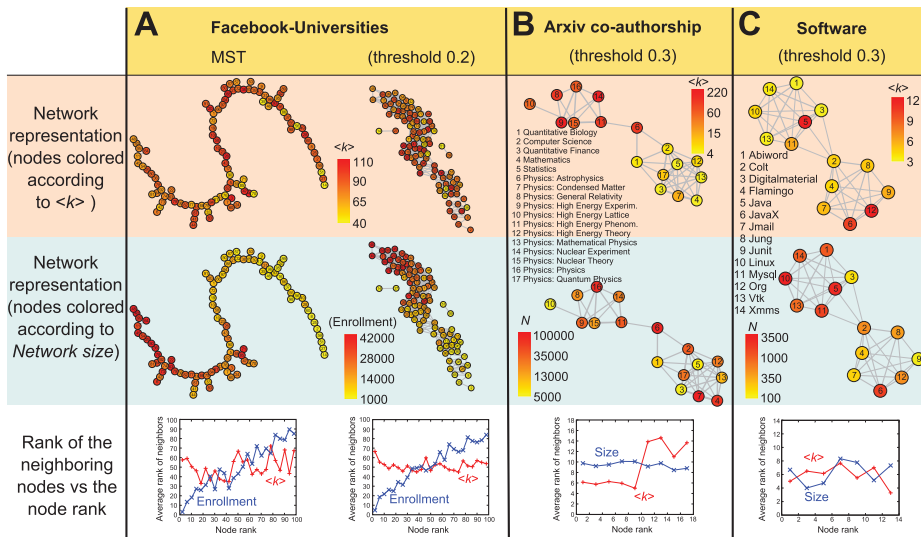


Fig. 3: (Color online) Comparison of static networks. (A) Minimum Spanning Tree and threshold-based network representation of similarities in the networks of Facebook friendship in 100 US universities. The color of the university nodes corresponds to either the average degree or the university size, in terms of enrollment size. The enrollment size is the key property for clustering. The plot at the bottom row compares the rank of a university to its neighbors rank. The enrollment size has a very hierarchical structure where ranks of the same order connect to each other, in contrast to average degree ranking where a nodes rank cannot predict the rank of its neighbors. (B) The similarity network of scientific fields, based on co-authorship, exhibits the opposite trend. The average degree is a nice indicator for clustering, while the network size is not. This result is supported by the plot comparing the rank with the neighbors rank. (C) The network of similarity between software projects cannot be clustered according to either the average degree or the network size. The two modules correspond however to networks that were built by two independent methods.

important factor in this case is the average degree of an author, *i.e.* fields with large number of co-authors yield similar networks with each other. This classification of networks according to an underlying structural property does not trivially result from the n -tangle method. In the example of software project networks in fig. 3(C) we were not able to determine any particular structural property that separates the projects in the n -tangle networks. Interestingly, each of the two modules in fig. 3(C) includes software projects that were generated by different circumstances. This method therefore, not only allows comparison across networks, but enables hypothesis testing about which facets might be the most salient organizational features that drive the emergence of networks within the systems studied.

We also applied this method to characterize network evolution. In the examples of the Internet growth [22] and online social-networking evolution [23] in fig. 4, we compare the network at a given time with the same network at subsequent times. The starting date for the Internet data was January 2004. Our method indicates that the Internet topology was already fixed in time by January 2004 and did not change much by November 2007, when the network had already doubled in size. This result holds across all subgraph sizes, and is also consistent with the macroscopic fact that the average link density was declining slowly from $2.4 \cdot 10^{-4}$ to $1.4 \cdot 10^{-4}$ over three years. On the contrary, the Facebook-like online network shows a stable behavior only at small scales n . The number of

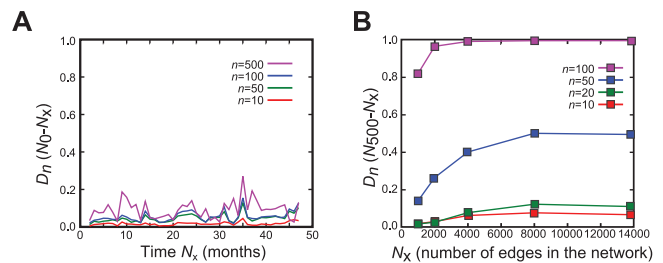


Fig. 4: (Color online) Comparison of network evolution. (A) Similarity of the Internet at the AS level with itself as a function of time. We compare the KS index $D_n(N_x - N_0)$ of the Internet topology at $N_0 =$ January 2004 with the topology at time N_x , which is increasing monthly. Independently of the scale n , the topology remains the same throughout the network evolution for 3 years. (B) Comparison of the KS index in social networking friendships as a function of time. We compare the network topology of the early network containing 500 links with the networks at subsequent times. The network remains the same for small values of n , but changes drastically at larger scales.

edges in the network increases by a factor of 25, but the n -tangle density remains very similar at any time when $n < 20$. When we consider larger n values, though, there is a very sharp change between the initial reference network and the subsequent instances of it. Therefore, within the same network the small-scale structures remain the same, while larger-scale structures evolve into different forms. The method can therefore separate structurally stabilized

networks over time from unstabilized ones. Moreover, in networks of evolving topology we can identify differences in the stability of short-scale and larger-scale structures. This may therefore enable accurate estimation of the quality of approximation of static snapshots of continually shifting networks, which has been shown to be of critical importance in areas such as epidemiology [24].

The calculation of the n -tangle density provides a simple and powerful method for efficient network comparison. It can also be used to compare the structure of multilayer networks [25], either among the layers of a network or across multilayer networks. Understanding the degree of similarity between two networks is the key to promote the classification of networks into clusters for further analysis of their common features that would otherwise remain unknown, and allows us to hypothesize meaningfully about how these clusters may capture fundamental properties of networks and the systems they represent.

* * *

We thank the Department of Homeland Security for funds in support of this research through the CCICADA Center at Rutgers, and NSF EaSM grant No. 1049088. We are grateful to the following people, for making their animal network data available to us: C. BERMAN, J. R. MADDEN, J. WOLF, D. B. McDONALD, P. C. CROSS, and S. SUNDARESAN.

Appendix: datasets. – The datasets we used in this study and their sources are as follows:

Animal affiliation networks. We have compiled a set of 37 empirically determined social networks in different species that can be found in the published literature.

Facebook in 100 universities. These networks are based on Facebook friendship connections in 100 colleges and universities in USA on September 2005. The data have been made publically available (<http://people.maths.ox.ac.uk/~porterm/data/facebook100.zip>) and have been analyzed in ref. [18].

Arxiv co-authorship networks. We downloaded the entire database of all papers in arxiv.org from 1991 until December 31, 2012. We used the sites classification of papers into 18 broad categories, and created one network for each category using all the papers in that field, e.g. Mathematical Physics, Computer Science, Math, etc. The networks connect co-authors of a paper in each category.

Software networks. We used networks of software from two sources: 1) The data for junit, jmail, flamingo, jung, colt, org, java, and javax were downloaded from <http://lovro.lpt.fri.uni-lj.si/publications.jsp?show=ssc>. The analysis of these data was done in [21]. 2) We also used the software packages Abiword, DigitalMaterial, Linux, Mysql, VTK, and XMMS from ref. [20].

Internet (evolving network). We downloaded the CAIDA Autonomous System graphs from January 2004 to November 2007 from the SNAP Stanford datasets in

<http://snap.stanford.edu/data/as-caida.html>. The data are described and analyzed in ref. [22].

Messages in an online social networking site. The dataset was downloaded from http://toreopsahl.com/datasets/#online_social_network and has been analyzed in [23]. It includes online messages sent among students at the University of California, Irvine, through a Facebook-like Social Network. Each message was time-stamped so we were able to follow the entire network evolution. Our starting point was when the first 500 links were created, and we sampled snapshots of the network with 1000, 2000, 4000, 8000, and 13838 links.

Gnutella sharing. We used the 9 snapshots of a peer-to-peer Gnutella network [26], where nodes represent hosts and links are the connections between these hosts.

Protein interaction networks. We used 9 protein interaction networks from BioGrid [27] for the following species: *A. thaliana*, *C. elegans*, *D. melanogaster*, *H. sapiens*, *M. musculus*, *P. falciparum*, *R. norvegicus*, *S. cerevisiae*, and *S. pombe*.

Metabolic networks. We used the 43 metabolic networks from ref. [28].

Road networks. A node in this network represents an intersection and the links correspond to the roads that connect these intersections. The three state-wide networks we used were for California, Pennsylvania, and Texas [29].

Thesauri networks. We extracted the networks from 5 thesauri datasets, where nodes represent words and the links indicate that the two words are synonyms. These data were extracted from the LibreOffice Thesaurus for English (UK), English (US), Spanish (AR), Spanish (ES), and Spanish (VE).

Web networks. The datasets for the web networks of Berkeley and Stanford, Google, Notre Dame, and Stanford [29] were downloaded from the Stanford SNAP database.

Amazon co-purchase. This network connects items that were frequently purchased together in amazon.com, as found by crawling software [30]. The data were downloaded from the Stanford SNAP database.

REFERENCES

- [1] BARABÁSI A.-L., *Nat. Phys.*, **8** (2012) 14.
- [2] ALBERT R. and BARABÁSI A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.
- [3] WOESE C. R., KANDLER O. and WHEELIS M. L., *Proc. Natl. Acad. Sci. U.S.A.*, **87** (1990) 4576.
- [4] MATTIUSI V., TUMMINELLO M., IORI G. and MANTEGNA R. N., *Comparing correlation matrix estimators via Kullback-Leibler divergence*, available at SSRN: ssrn.com/abstract=1966714.
- [5] MADI A. *et al.*, *Chaos*, **21** (2011) 6109.
- [6] KENETT D. Y., PREIS T., GUR-GERSHGORIN G. and BEN-JACOB E., *Int. J. Bifurcat. Chaos*, **22** (2012) 1250181.
- [7] KENETT D. Y., HUANG X., VODENSKA I., HAVLIN S. and STANLEY H. E., preprint arXiv:1402.1405 (2014).

- [8] MILO R., SHEN-ORR S., ITZKOVITZ S., KASHTAN N., CHKLOVSKII D. and ALON U., *Science*, **298** (2002) 824.
- [9] PRZULJ N., *Bioinformatics*, **23** (2007) e177.
- [10] BASKERVILLE K. and PACZUSKI M., *Phys. Rev. E*, **74** (2006) 051903.
- [11] ONNELA J.-P., FENN D. J., REID S., PORTER M. A., MUCHA P. J., FRICKER M. D. and JONES N. S., *Phys. Rev. E*, **86** (2012) 036104.
- [12] PAPADIMITRIOU P., DASDAN A. and GARCIA-MOLINA H., *J. Internet Serv. Appl.*, **1** (2010) 19.
- [13] STEPHENS M. A., *J. Am. Stat. Assoc.*, **69** (1974) 730.
- [14] UGANDER J., BACKSTROM L. and KLEINBERG J., in *Proceedings of the 22nd International World Wide Web Conference* (ACM, New York, NY) 2013, p. 1307.
- [15] MICHENER C. D. and SOKAL R. R., *Evolution*, **11** (1957) 130.
- [16] ALEXANDER R. D., *Ann. Rev. Ecol. Syst.*, **5** (1974) 325.
- [17] McDONALD D. B., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 10910.
- [18] TRAUD A. L., MUCHA P. J., PORTER M. A., *Physica A*, **391** (2012) 4165.
- [19] NEWMAN M. E. J., *Proc. Natl. Acad. Sci. U.S.A.*, **98** (2001) 404.
- [20] MYERS C. R., *Phys. Rev. E*, **68** (2003) 046116.
- [21] SUBELJ L. and BAJEC M., *Physica A*, **390** (2011) 2968.
- [22] LESKOVEC J., KLEINBERG J. and FALOUTSOS C., in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (ACM, New York, NY) 2005, p. 177.
- [23] OPSAHL T. and PANZARASA P., *Soc. Netw.*, **31** (2009) 155.
- [24] FEFFERMAN N. H. and NG K. L., *Phys. Rev. E*, **76** (2007) 031919.
- [25] BULDYREV S. V., PARSHANI R., PAUL G., STANLEY H. E. and HAVLIN S., *Nature*, **464** (2010) 08932.
- [26] LESKOVEC J., KLEINBERG J. and FALOUTSOS C., *ACM Trans. Knowl. Discov. Data*, **1** (2007) 2.
- [27] STARK C., BREITKREUTZ B. J., REGULY T., BOUCHER L., BREITKREUTZ A. and TYERS M., *Nucleic Acids Res.*, **34** (2006) D535.
- [28] JEONG H., TOMBOR B., ALBERT R., OLTVAI Z. N. and BARABÁSI A.-L., *Nature*, **407** (2000) 651.
- [29] LESKOVEC J., LANG K. J., DASGUPTA A. and MAHONEY M., *Internet Math.*, **6** (2009) 29.
- [30] LESKOVEC J., ADAMIC L. A. and HUBERMAN B. A., *ACM Trans. Web*, **1** (2007) 5.
- [31] CROSS P. C., LLOYD-SMITH J. O., BOWERS J. A., HAY C. T., HOFMEYR M. and GETZ W. M., *Ann. Zool. Fenn.*, **41** (2004) 879.
- [32] DE SILVA S. and WITTEMYER G., *Int. J. Primatol.*, **33** (2012) 1125.
- [33] CHAVERRI G., *Behav. Ecol. Sociobiol.*, **64** (2010) 1619.
- [34] CORNER L. A. L., PFEIFFER D. U. and MORRIS R. S., *Prev. Vet. Med.*, **59** (2003) 147.
- [35] MURRAY C. M., *Int. J. Primatol.*, **28** (2007) 853.
- [36] SCHURCH R., ROTHENBERGER S. and HEG D., *Philos. Trans. R. Soc. B*, **365** (2010) 4089.
- [37] MANNO T. G., *Anim. Behav.*, **75** (2008) 1221.
- [38] RUTZ C., BURNS Z. T., JAMES R., ISMAR S. M. H., BURT J., OTIS B., BOWEN J. and ST CLAIR J. J. H., *Curr. Biol.*, **22** (2012) R669.
- [39] LUSSEAU D., SCHNEIDER K., BOISSEAU O. J., HAASE P., SLOOTEN E. and DAWSON S. M., *Behav. Ecol. Sociobiol.*, **54** (2003) 396.
- [40] FORMICA V. A., WOOD C. W., LARSEN W. B., BUTTERFIELD R. E., AUGAT M. E., HOUGEN H. Y. and BRODIE E. D. III, *J. Evol. Biol.*, **25** (2012) 130.
- [41] SHORROCKS B. and CROFT D. P., *Afr. J. Ecol.*, **47** (2009) 374.
- [42] PSORAKIS I., ROBERTS S. J., REZEK I. and SHELDON B. C., *J. R. Soc. Interface*, **9** (2012) 3055.
- [43] CANTOR M., WEDEKIN L. L., GUIMARAES P. R., DAURA-JORGE F. G., ROSSI-SANTOS M. R. and SIMOES-LOPES P. C., *Anim. Behav.*, **84** (2012) 641.
- [44] CROFT D. P., KRAUSE J., DARDEN S. K., RAMNARINE I. W., FARIA J. J. and JAMES R., *Behav. Ecol. Sociobiol.*, **63** (2009) 1495.
- [45] HOLEKAMP K. E., SMITH J. E., STRELIOFF C. C., VAN HORN R. C. and WATTS H. E., *Mol. Ecol.*, **21** (2012) 613.
- [46] GODFREY S. S., BULL C. M., JAMES R. and MURRAY K., *Behav. Ecol. Sociobiol.*, **63** (2009) 1045.
- [47] McDONALD D. B., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 10910.
- [48] BLUMSTEIN D. T., WEY T. W. and TANG K., *Proc. R. Soc. B*, **276** (2009) 3007.
- [49] SUNDARESAN S. R., FISCHHOFF I. R., DUSHOFF J. and RUBENSTEIN D. I., *Oecologia*, **151** (2007) 140.
- [50] GUIMARAES P. R., DE MENEZES M. A., BAIRD R. W., LUSSEAU D., GUIMARAES P. and DOS REIS S. F., *Phys. Rev. E*, **76** (2007) 042901.
- [51] FLACK J. C., GIRVAN M., DE WAAL F. B. M. and KRAKAUER D. C., *Nature*, **439** (2006) 426.
- [52] MCSWEENEY D. J., BAIRD R. W., MAHAFFY S. D., WEBSTER D. L. and SCHORR G. S., *Mar. Mamm. Sci.*, **25** (2009) 557.
- [53] CROFT D. P., JAMES R. and KRAUSE J., *Exploring Animal Social Networks* (Princeton University Press) 2008.
- [54] BERMAN C. M., RASMUSSEN K. L. R. and SUOMI S. J., *Anim. Behav.*, **53** (1997) 405.
- [55] WOLF J. B. W. and TRILLMICH F., *Proc. Biol. Sci.*, **275** (2008) 2063.
- [56] ZHANG P., LI B., QI X., MACINTOSH A. J. J. and WATANABE K., *Int. J. Primatol.*, **33** (2012) 1081.
- [57] TEMPLETON C. N., REED V. A., CAMPBELL S. E. and BEECHER M. D., *Behav. Ecol.*, **23** (2012) 141.
- [58] HAMEDE R. K., BASHFORD J., MCCALLUM H. and JONES M., *Ecol. Lett.*, **12** (2009) 1147.
- [59] SUGAWARA K., *Primates*, **20** (1979) 21.
- [60] MADDEN J. R., DREWE J. A., PEARCE G. P. and CLUTTON-BROCK T. H., *Behav. Ecol. Sociobiol.*, **64** (2009) 81.
- [61] RYDER T. B., McDONALD D. B., BLAKE J. G., PARKER P. G. and LOISELLE B. A., *Proc. R. Soc. B*, **275** (2008) 1367.