Precision as a measure of predictability of missing links in real networks

Guillermo García-Pérez,[1, 2, *] Roya Aliakbarisani,[3, 4, *] Abdorasoul Ghasemi,[4] and M. Ángeles Serrano[3, 5, 6]

[1]*QTF Centre of Excellence, Turku Centre for Quantum Physics,*
Department of Physics and Astronomy, University of Turku, FI-20014 Turun Yliopisto, Finland
[2]*Complex Systems Research Group, Department of Mathematics and Statistics,*
University of Turku, FI-20014 Turun Yliopisto, Finland
[3]*Departament de Física de la Matèria Condensada,*
Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
[4]*Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran 1631714191, Iran*
[5]*Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain*
[6]*ICREA, Pg. Lluís Companys 23, E-08010 Barcelona, Spain*

Predicting missing links in real networks is an important open problem in network science to which considerable efforts have been devoted, giving as a result a vast plethora of link prediction methods in the literature. In this work, we take a different point of view on the problem and focus on predictability instead of prediction. By considering ensembles defined by well-known network models, we prove analytically that even the best possible link prediction method, given by the ensemble connection probabilities, yields a limited precision that depends quantitatively on the topological properties—such as degree heterogeneity, clustering, and community structure—of the ensemble. This suggests an absolute limitation to the predictability of missing links in real networks, due to the irreducible uncertainty arising from the random nature of link formation processes. We show that a predictability limit can be estimated in real networks, and we propose a method to approximate such bound from real-world networks with missing links. The predictability limit gives a benchmark to gauge the quality of link prediction methods in real networks.

[] J. Slingo and T. Palmer, Phil Trans R Soc A **369**, 4751 (2011).

[] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley, Proc Natl Acad Sci USA **112**, 2325 (2015).

[] L. Lü and T. Zhou, Physica A **390**, 1150 (2011).

[] P. Erdös and P. Rényi, Publ Math Debrecen **6**, 290 (1959).

[] J. Park and M. E. J. Newman, Phys Rev E **68**, 026112 (2003).

[] M. Á. Serrano, D. Krioukov, and M. Boguñá, Phys Rev Lett **100**, 078701 (2008).

[] B. Karrer and M. E. J. Newman, Phys Rev E **83**, 016107 (2011).

[] M. E. J. Newman, Phys Rev E **64**, 025102 (2001).

[] L. A. Adamic and E. Adar, Soc Networks **25**, 211 (2003).

[] T. Zhou, L. Lü, and Y.-C. Zhang, Eur Phys J B **71**, 623 (2009).

[] A. Muscoloni and C. V. Cannistraci, arXiv:1707.09496. Preprint, posted Jul 29, 2017 (2017).

[] Z. Liu, J.-L. He, K. Kapoor, and J. Srivastava, PloS one **8**, e72908 (2013).

[] M. Boguñá, F. Papadopoulos, and D. Krioukov, Nat Commun **1**, 62 (2010).

[] F. Papadopoulos, R. Aldecoa, and D. Krioukov, Phys Rev E **92**, 022807 (2015).

[] M. A. Riolo, G. T. Cantwell, G. Reinert, and M. E. J. Newman, Phys Rev E **96**, 032310 (2017).

[] T. Vallès-Català, T. P. Peixoto, M. Sales-Pardo, and R. Guimerà, Phys Rev E **97**, 062316 (2018).

[] M. Á. Serrano, M. Boguná, and A. Vespignani, Proc Natl Acad Sci USA **106**, 6483 (2009).

* These two authors contributed equally