

The Analysis of the Patent Citation Network

Péter Érdi

1: Center for Complex System Studies, Kalamazoo College, Kalamazoo, MI

2: Wigner Research Centre for Physics, Hungarian Academy of Sciences, Budapest, Hungary

0. The patent citation network, a huge, growing directed graph, is the result of a social game played by governmental institutions, universities, individual inventors, private firms, research institutes, patent lawyers, patent examiners and attorneys. Given that the citation network encapsulates information about technological relationships and progress provided by those players, understanding its development could help to inform policy makers as to how to allocate resources optimally to research and development. Some analysis of the USPTO and NBER databases at the level of individual patents, and of "patent clusters" will be demonstrated.

1. *The Rules Behind the Development of the Patent Citation Network*

The likelihood that a patent will be cited, was found to be approximately separable into a product of a two one-variable functions, one variable is the number of citations already received (in-degree) and the other is the age measured in patent number units. The age-dependence displays a peak at a low age value and a long power law tail, suggesting that some patented technologies have very long-term effects. The in-degree dependence exhibits super-linear preferential attachment. The preferential attachment exponent has been increasing since 1991, suggesting that patent citations are increasingly concentrated on a relatively small number of patents. The overall average probability that a new patent will be cited by a given patent has increased slightly during the same period. Some possible implications of our results for patent policy is discussed.

2. *Prediction of Emerging Technologies*

The central element of predictive analytics is the predictor, a variable that can be defined for an individual, organization or other entity and employed to predict its future behavior. Here we define a citation vector for each patent to play the role of a predictor, i.e., to characterize the temporal change of technological fields. Each coordinate of the citation vector represents how frequently the patent has been cited by other patents in a particular technological category. Changes in this citation vector over time reflect the changing role that a particular patented technology is playing as a contributor to later technological development.

To track the development of technological clusters, we employ clustering algorithms based on a measure of similarity defined using the citation vectors. We hypothesize that patents with similar citation vectors will belong to the same technological field. Thus, the formation over time of new clusters should correspond to the emergence of new technological directions. Using this approach we show how past changes in the clusters can be detected, and future changes can be predicted.

Acknowledgments

Collaboration Kinga Makovi (Columbia Univ), Katherine Strandburg (NYU), Jan Tobochnik (Kalamazoo College) and with members of the *Budapest Complex Systems and Computational Neuroscience Group* at Wigner Research Centre for Physics (Zoltán Somogyvári, Péter Volf, László Zalányi) is acknowledged. Thanks to the Henry R. Luce Foundation for general support.