Innovation Networks: Review

Alun Meredith

1 Introduction

The field of scale-free networks is relatively young with the bulk of research since 1999 [4]. A defining trait of scale-free networks is to have simple interactions creating emergent behaviour in a way in which it is hard to predict. Because of this simulations and modelling has proven a powerful tool in understanding the links between the two.

Patent Networks have been studied in many ways, from natural language processing [2], graph theoretical approaches and exploring through evolutionary algorithms.

One of the biggest challenges is the difficulty measuring the value of the patent and there is a lot of debate about the efficacy of different measures. To this end different characteristics are often used to build the network: technology codes, text similarity and primarily bibliographic citations.

2 Graph Theory

The foundation of the graph theoretical approach is Price's model which [5]explains the Mathew effect found in citation rates for patents [6] by modelling the attachment of new nodes as a linear function of the number of edges of that node. There has been a wealth of research improving the model with concepts of non-linear growth, local events, growth constraints and initial attractiveness [7].

In 2007 Csardi et al. published the first paper examining the US patent office database from this perspective [8]. They did this by by applying a basic model that assumes the attractiveness of a node (rate at which new nodes attach to it) is a function of the age and number of citations of the node. Normalising for the growth of the patent network over time they found the total attractiveness of the system over time could only be replicated with a super-linear preferential attachment model.

They also explore the idea that an increase in the number of citations per patent over time has been coupled with a fundamental change in the structure of the network. Finding that the level of stratification starts to increase in alignment with the higher citation rates.

Valverde et. al. [9] builds on this work by suggesting a form for the extended power law and shows with correct parameters this form can describe both the scale-invariant and exponential tail of the citation count distribution. They also explore the clustering and modularity of the system. The clustering coefficient being approximately inversely proportional to the number of citations, suggesting a hierarchical structure.

There is an increasing body of research in academic bibliographic networks which aims to explain more of the mechanics behind citation formation in their graph network models. This includes the propagation of errors which suggests 70 - 80% of citations are copy and pasted from a secondary source [10] or the study of redundant edges to show that the majority of references (70%) are secondary [11].

A recent study builds on this theme exposing the weakness of relying on simple models of patent networks [1]. Bernard Gress (2009) investigated the ratios between citations given and received in different technology groups. High number and diversity of citations given was treated as an indication of generality and number of citations received of productivity and originality. He then compared these measures and how they varied over time for different technology categories. He primarily concludes that these categories are fundamentally different therefore research needs to take this into account.

Byungun Yoon [3] built a patent network from weighted term similarities of patent documents rather than bibliometric citations; using standard bag of words methods and comparing these measures to analogies in citation networks. Through inspection they argue that the centrality of their network yields a more relevant approximation of impact because it is less biased by age and preferential attachment mechanisms.

3 Innovation Evolution

There is a body of research exploring the analogy between the evolution of innovation and biological evolution. The premise is that each invention is built from the recombination of previous inventions. The two models have their differences, for example there is a limited concept of 'death' in innovation as very old patents may still be cited by new ones and it is hard to think of bibliometric patent networks as direct lineages.

Yeoun et al. [12] explores this idea of invention as a recombination process by looking at the use of technology codes in patents as a proxy for novelty. Technology codes map the technological niche of a patent into categories and subcategories. Patents can have combinations of technology codes. They show that as the number of patents increases the number of new codes being generated falls off while new code combinations maintains a power-law, concluding new technologies has a minimal role relative to recombination. They also show that 40% of patents use existing combinations vs. new ones suggesting these are incremental improvements.

Technology code combination distributions do not age in the same way as bibliometric citations, codes appear not to age with 99% of codes being used at least once every 7 years.

They also look at the dissimilarity of the codes as a proxy for novelty. If a patent is used in a very different field from its parents it is argued that it is more likely to be a bigger leap in novelty. Categorising patents as either narrow or broad and using count as a metric, we only get a sense how novelty has changed with time and not any of the network factors which may be present here, such as the distribution of novelty could be a power law or the degree of novelty can be a measure of linkage between clusters.

The limitations of the evolutionary analogy are loosely addressed warning that citations in patents aren't directly related to lineage but about carving a legal niche and there being no good metric of fitness for patents.

Buchanan et al. [13] glosses over some of these limitations, using the number of citations a patent has as an "impact" metric, a proxy for fitness. Prior art citations also function as a proxy for combinatorial lineage. They tell the story of the most cited patents in the network over the past 30 years and show that such a distribution of citations cannot come from random natural selection and therefore must be due to adaptive selection in an evolutionary model of innovation. This argument is an evolutionary perspective on the random network vs. preferential attachment network di'fferentiation.

They focus on showing this idea more robustly incorporating a multitude of normalisation techniques and simulating a null hypothesis random network model by sampling existing data, rather than building a clean model from scratch. Observing the familiar hallmarks of a fat-tailed distribution they conclude that these "superstars" high impact is due to adaptive features, however they do not address the role of preferential attachment here, how many of the citations received are due to 'rich getting richer' mechanics or due to the intrinsic quality of those innovations.

Their paper also investigates the dissimilarity of technology codes as a proxy for novelty making the claim that large leaps in novelty are responsible for the largest "impact" patents. However because its scope is only looking at the 20 most cited patents falls short of being able to make such a general argument about the network as a whole.

Finally Arthur et al. [14] makes the most direct contrast between the search process of a genetic algorithm and the evolution of technology. In their paper they simulate an evolving population of logical circuits starting from simple logic gates in order to meet a selection of logical needs. Analysing the evolution of the population and resulting network.

They find many of the typical evolutionary features present such as building blocks being formed as intermediary steps to complex solutions i.e. the building block hypothesis. Sub-optimal solutions slowly become extinct after better solutions emerge.

Complex features are also observed such as a loose power law distribution of edges in the network and avalanches of redundancy as new technologies replace old ones and their dependencies, the size of these redundancies follows a power law showing self-organised criticality.

The paper incorporates a standard genetic algorithm; ignoring many of the observed differences between natural selection and the evolution of innovation, such as holding a finite population therefore incorporating the ”death” of patents as they are replaced, and use of random selection. Both of these were shown earlier to not be accurate in the patent network, despite this they achieve results similar to observed patent networks, further research could conclude that many of the differences observed naturally arrive from a simple model for example patent ageing could be a result from the saturation of combinatorial space around older technologies.

4 Conclusion

There are three main directions in which Patent Networks are studied, as a branch of already existing Price model networks, through natural language processing techniques and through an analogy to evolutionary processes.

We have seen how efforts have been made to link the evolution of innovations to a Darwinian evolutionary models, that despite some success a lot more effort needs to be done to incorporate the differences between natural and technological evolution into models. linking these more strongly with networks models can be a key to understanding the mechanics of the innovation of evolution.

There have only been a few studies of this dataset from a graph theoretical perspective. There is room for more work in this area especially when paralleling the advances of the academic citation literature to include more subtlety to the models.

The greatest challenge to studying patent networks lies in the efficacies of citations as a measure of impact, there has been great debate in the papers referenced here to what extent impact can be measured this way and to what degree these measures can be improved.

References

[1] B. Gress, “Properties of the uspto patent citation network: 1963–2002,” *World Patent Information*, vol. 32, no. 1, pp. 3–21, 2010.

[2] A. Abbas, L. Zhang, and S. U. Khan, “A literature review on the state-of-the-art in patent analysis,” *World Patent Information*, vol. 37, pp. 3–13, 2014.

[3] B. Yoon and Y. Park, “A text-mining-based patent network: Analytical tool for high-technology trend,” *The Journal of High Technology Management Research*, vol. 15, no. 1, pp. 37–50, 2004.

[4] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[5] D. d. S. Price, “A general theory of bibliometric and other cumulative advantage processes,” *Journal of the American society for Information science*, vol. 27, no. 5, pp. 292–306, 1976.

[6] R. K. Merton *et al.*, “The matthew effect in science,” *Science*, vol. 159, no. 3810, pp. 56–63, 1968.

[7] R. Albert and A.-L. Barabasi, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, vol. 74, pp. 47–97, 2002.

[8] G. Csárdi, K. J. Strandburg, L. Zalányi, J. Tobochnik, and P. Érdi, “Modeling innovation by a kinetic description of the patent citation system,” *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 2, pp. 783–793, 2007.

[9] S. Valverde, R. V. Solé, M. A. Bedau, and N. Packard, “Topology and evolution of technology innovation networks,” *Physical Review E*, vol. 76, no. 5, p. 056118, 2007.

[10] M. V. Simkin and V. P. Roychowdhury, “Stochastic modeling of citation slips,” *Scientometrics*, vol. 62, no. 3, pp. 367–384, 2005.

[11] J. R. Clough, J. Gollings, T. V. Loach, and T. S. Evans, “Transitive reduction of citation networks,” *Journal of Complex Networks*, vol. 3, no. 2, pp. 189–203, 2015.

[12] H. Youn, D. Strumsky, L. M. Bettencourt, and J. Lobo, “Invention as a combinatorial process: evidence from us patents,” *Journal of The Royal Society Interface*, vol. 12, no. 106, p. 20150272, 2015.

[13] A. Buchanan, N. H. Packard, and M. A. Bedau, “Measuring the evolution of the drivers of technological innovation in the patent record,” *Artificial life*, vol. 17, no. 2, pp. 109–122, 2011.

[14] W. B. Arthur and W. Polak, “The evolution oftechnology within a simple computer model,” *Complexity and the Economy*, 2014.