Network Science and Law: A Sales Pitch and an Application to the “Patent Explosion”

Katherine J. Strandburg
DePaul University College of Law

Gabor Csardi, Laszlo Zalanyi,
Jan Tobochnik, Peter Erdi
Kalamazoo College, KFKI Institute, Budapest
OUTLINE

I. NETWORK SCIENCE AND LAW: IS THERE A CONNECTION?

II. THE PATENT CITATION NETWORK

III. CITATION NETWORK GROWTH

IV. DEGREE DISTRIBUTION
WHY NETWORK SCIENCE IS RELEVANT TO LAW

- Networks are everywhere in law
  - Legal citations, patent citations, social networks, collaborative networks, networks of firms, etc.
- Legal scholarship is (to some extent) “applied social science”
  - seeks to understand and predict collective, social results of changes in legal rules, social realities
  - has mostly applied economics
  - generalize from pairwise interactions (>2000 cites to Prisoner’s Dilemma)
  - “mean field” approach (social norms)
WHY NETWORK SCIENCE IS RELEVANT TO LAW

• Network science provides analysis and modeling that accounts for heterogeneity, relatedness, local context in social structures and results could have implications for legal policy.

• Law provides “natural experiments” for network science (how does legal change affect network structure, collective behavior on network?)

• Law provides some very extensive datasets.
PATENT CITATION NETWORK

WHY STUDY IT?

- Very large network (> 4 million nodes)
- Data available electronically (NBER dataset + USPTO)
- Complete history available
- Comparison to scientific citation networks
- Relevance to patent policy
POLICY MOTIVATION

• Number of patents and patent applications is increasing exponentially over time, “explosion” since early 1980’s
• Perception that patents are increasingly “trivial”, patent “thicket”
• Various theories about how changes to the patent system have affected patent “quality”
• NAS study, FTC study, Patent Reform Act

USE NETWORK APPROACH TO STUDY HOW THE PATENT SYSTEM IS CHANGING
EXPONENTIAL GROWTH IN PATENTING

Log Patents over Time

Year
Log of Number
1750 1800 1850 1900 1950 2000 2050
Log Apps  Log Issued
BEHIND THE PATENT “EXPLOSION”

- Increasing pace of technology?
- Increasing breadth of patented technology? (Scientific breakthroughs or legal broadening of patentable subject matter)
- Decreasing patentability standard?

HAS THERE REALLY BEEN A SUBSTANTIVE CHANGE?
PATENT CITATION PRIMER

- Patentable inventions are NEW, USEFUL, and NONOBlVIOUS
- Determined by comparison to PRIOR ART (patents, publications, prior uses)
- Citation if “material to patentability”
  - improvement
  - distinguish
- Citations come from
  - applicants (who need not search)
  - examiners (who must search)
PATENT NETWORK GROWTH

- Stochastic growth model:
  Assume \( P[e \text{ cites node } i] = A(k_i(t); l_i(t))/S(t) \)
  \( A(k,l) \) “attractiveness”
  \( k \) - number of previous citations (in-degree)
  \( l \) - age in patent numbers
  \( S(t) \) normalization factor = sum \( A(k(t),l(t)) \)
  \( t \) - time in patent numbers

- Extract \( A(k,l) \), \( S(t) \) numerically from the data using a self-consistent iterative method

- Assume \( A(k,l) \) time-independent for now
PATENT NETWORK GROWTH

- Extract $A(k,l)$, $S(t)$ numerically from the data using an iterative method
- Method is general for any network for which history is known
- Can extract $A(\bullet, \bullet, \bullet, \ldots)$ for variables of interest, e.g., degree, age, recent degree
  - simulate network evolution
  - compare to assess validity of model (ONGOING WORK)
PATENT NETWORK GROWTH

- We don’t assume a particular form for $A(k,l)$ or $S(t)$, we extract them from the data.

- We only account for US patents, so we may systematically underestimate technological relationships.

- We don’t assume a particular meaning for citations (no assumption of knowledge flow).

- Time in patent numbers takes out any simple dependence on pace of technology.
SOME RESULTS

- Old patents less likely to be cited (aging)
- Patents already cited more likely to be cited (preferential attachment)
- \[ A(k,l) = (\text{approx.}) A_k(k)A_l(l) \]
  - approximately separable for all \( k \) and \( l \)
$A_k \sim k^\alpha$, $\alpha = 1.201+/-.006$
- Allow $\alpha$ to vary with time and look at a sliding time window of 500,000 patents
  - $\alpha$ decreasing or flat until ~1993 and increased since INCREASING STRATIFICATION, NOT JUST MORE PATENTING
INTERPRETATION OF INCREASING STRATIFICATION

- Most “citable” patents are increasingly “more citable” than “least citable” patents. Why?

MAY RELATE TO LEGAL CHANGES

- Lowering patentability standard?
  - consistent with anecdotal evidence
  - possibly related to lowered standard for nonobviousness

- Increased patenting of upstream science?
AGE DEPENDENCE

A(k, l) has the same dependence on l (age) for all k (citations)
AGE DEPENDENCE

• Two “pieces” to the aging function
  
  - Citation probability peaks about about 200K patents (around 1-2 years or so) for all k, “typical” citations
  
  - BUT there is a significant power law tail of delayed citations
Long tail of delayed citations for all $k$
Roughly Universal for $k > \sim 5$, exponent $\sim 1.6$
AGE DEPENDENCE

- Slow decay means that some patents continue to be cited for very long times, “pioneer” patents
- Slow decay is there for all values of k
- Some “sleeper” patents that have never or rarely been cited will be cited after long periods of time
- Suggests unpredictability of technological progress
- There is no natural “cutoff” for patent influence
Approximately log-normal distribution
Evolves from more power-law-like
Aging effects
• Power law tail (better than citations received)
• Mechanism? (not preferential attachment)
• Very different from scientific citations (exponential?)
Patent Examiner v. Inventor Citations

- Tail dominated by inventor citations
- Very different from scientific citations
- Mechanism? Related to patentee valuation of inventions?
References

Physics version:

Modeling Innovation by a Kinetic Description of the Patent Citation System, www.arxiv.org/physics/0508132


Law version: