Evolution of Innovation in Patent Citation Networks

Ryan Steed

Departments of Economics & Computer Science

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

Introduction

- Economists, historians and business leaders generally agree that innovation is inextricably linked to continued prosperity and national competitiveness.
- Accordingly, nations sponsor research and craft legislation, such as intellectual property protection, to stimulate innovation.
- To justify this investment and assess its benefits, **is there a rigorous way to quantifiably measure innovation and its spread with currently available data**?

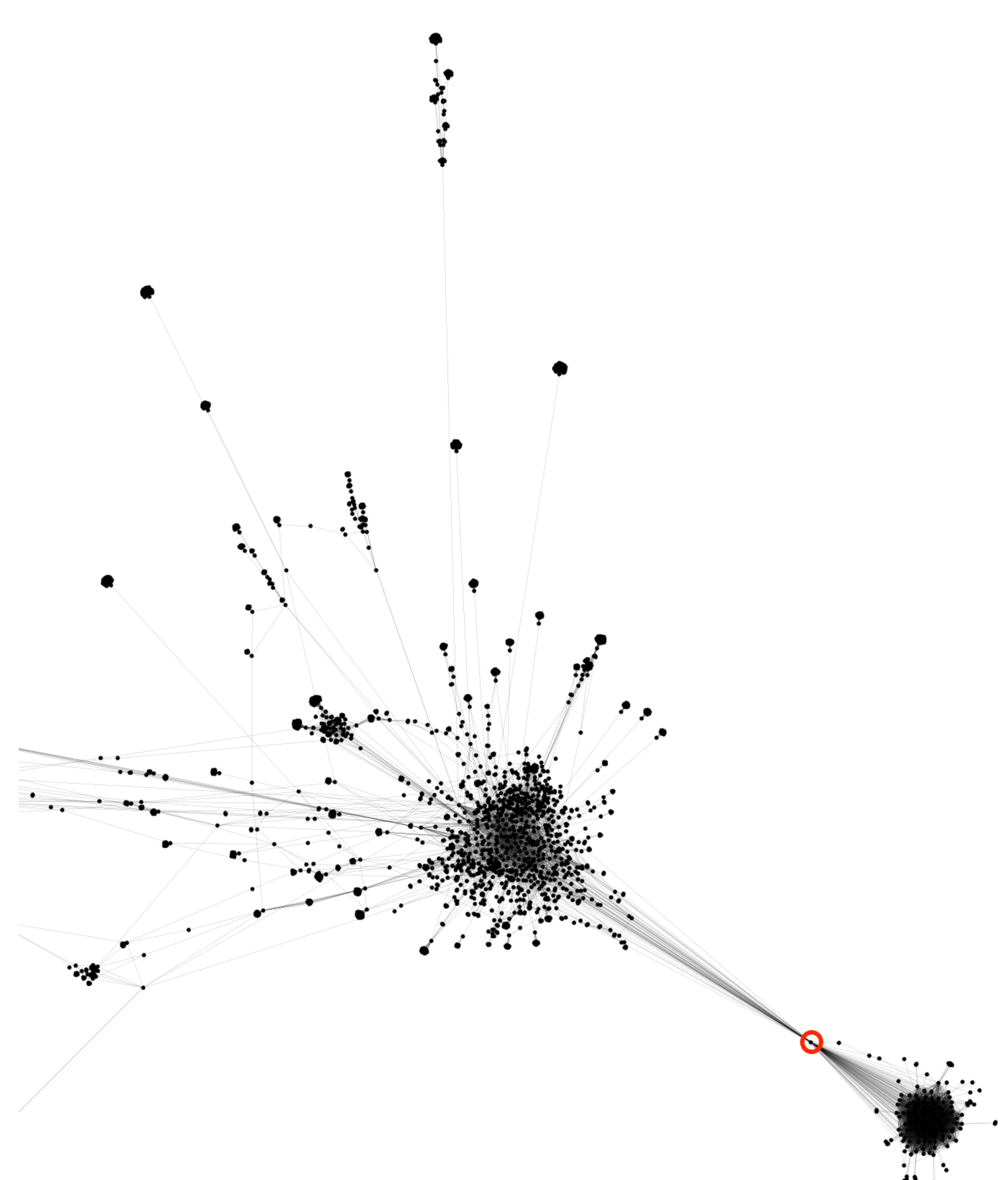

Figure 1: The citation network rooted on patent 3961197 (red) constitutes its descendant tree. Up to three generations are displayed.

Citation Network Data

- Patents comprise the best source of public IP data and compose a large network of patent nodes linked by citations, which represent knowledge flows.
- Networks are constructed from 133,760 patents granted 1976-2018 in several USPC tech sectors.
- TKC index is calculated for each patent in 20 week bins over the same period.
- Observable exogenous features are also collected, including the claims, inventor, and assignee.

Total Knowledge Contribution

How much original impact does a given patent have on future R & D? For patent i:

- TKC is total knowledge contribution;
- W is any topological measure of patent importance (e.g. the h-index, or out-degree centrality);
- b is the number of backward citations;
- n is the number of forward citations;
- the discount factor $0 < \lambda \leq 1$.

$$TKC_i = W_i + \sum_{j=1}^{n_i} \lambda \frac{K_j}{b_i}$$

- TKC is calculated by recursive traversal of the descendant tree (Fig. 1).
- TKC is robust to importance metric selection.
- TKC tends to be higher in older, more established technology sectors (Fig. 2).

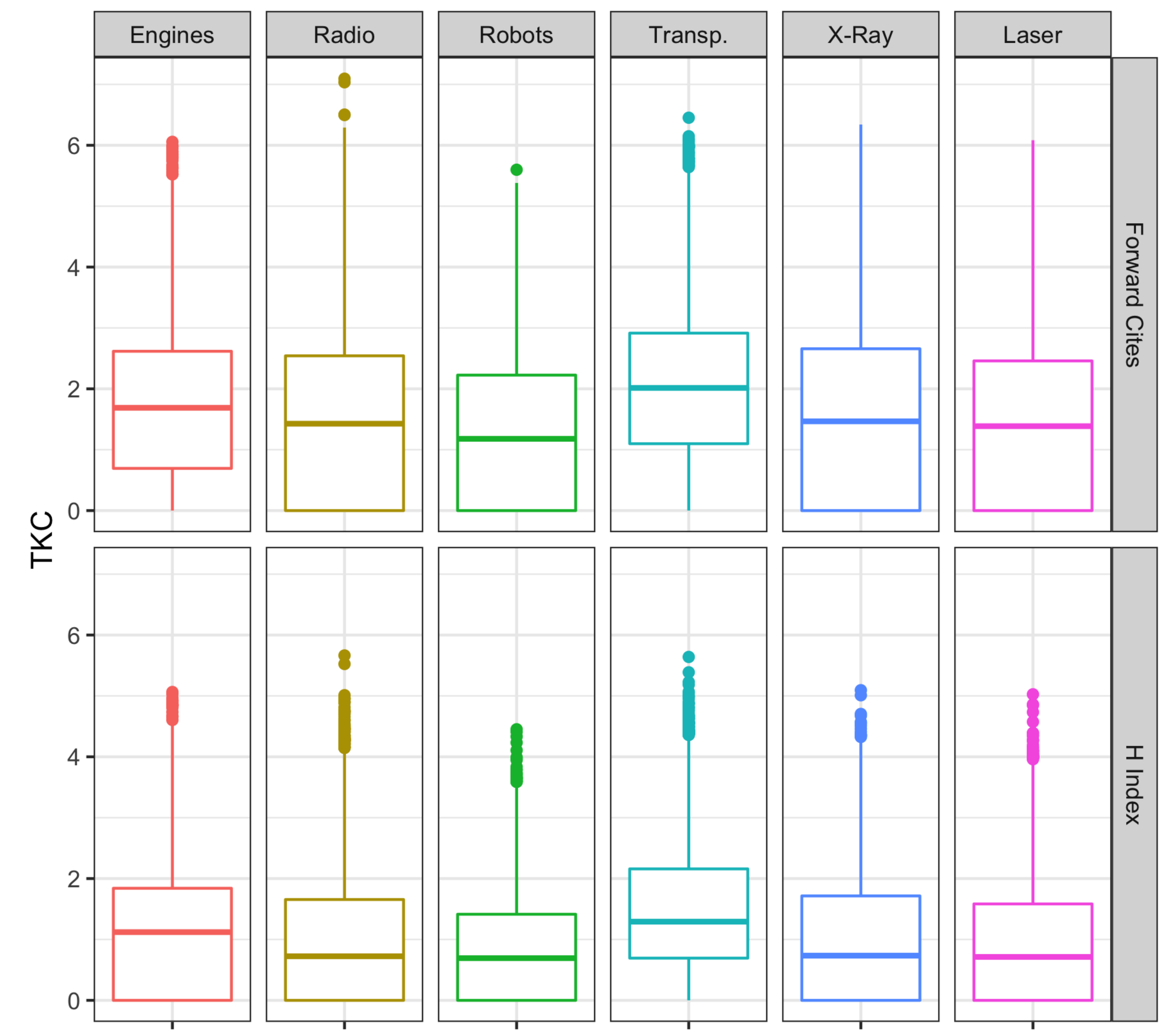

Figure 2: Boxplot distribution of $\log(TKC)$, calculated using two different weighting methods for each test sector. Distribution means are significantly different at the 1% level in a pairwise t-test.

Forecasting: A Case Study

The 2011 America Invents Act (AIA) replaced the first-to-invent policy with the European first-to-file system.

- An ARIMA time series forecast on pre-AIA mean TKC significantly differs from actual data (Fig. 3).
- Time dummy parameters estimated with a pooled OLS cross-sectional time series regression (controlling for exogenous patent features) steadily increase until just before the AIA is signed, then sharply decline (Fig. 4).

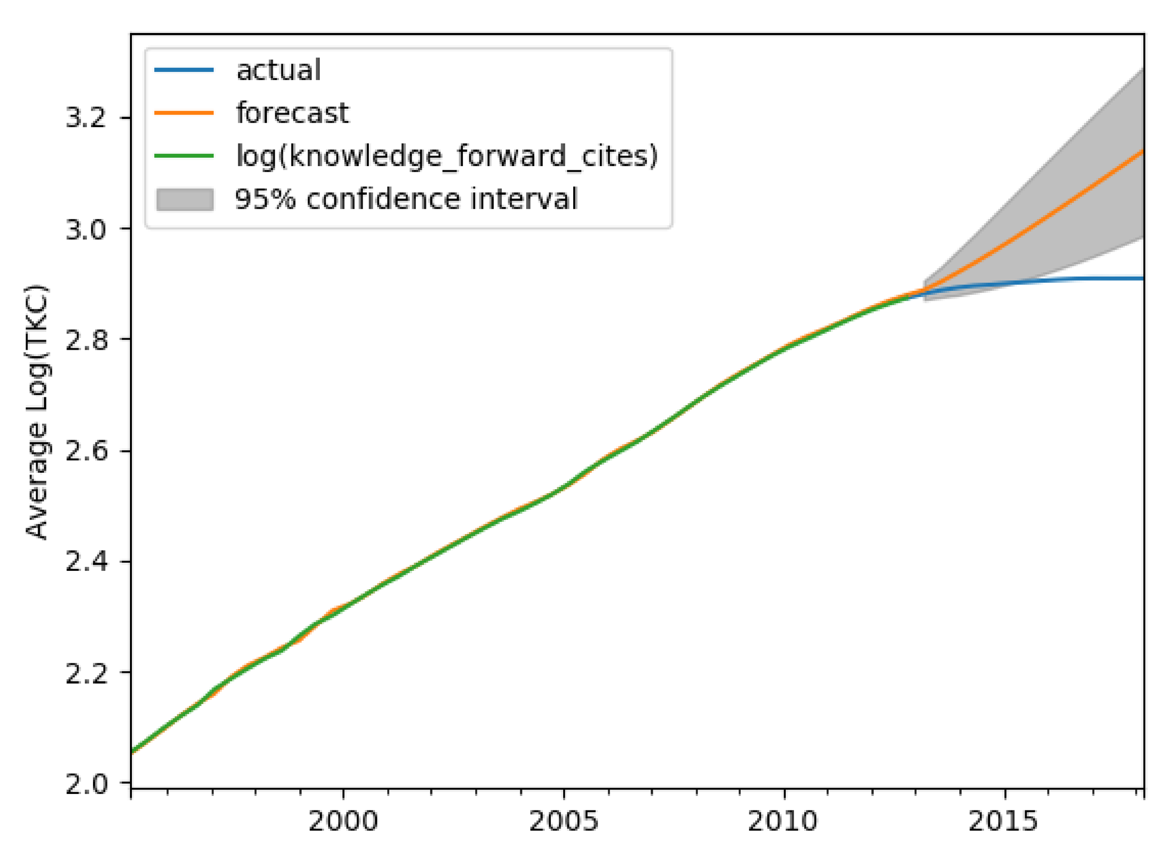

Figure 3: Average knowledge contribution ARIMA(2,1,0) forecast for all datasets, compared before and after the AIA effective date, with 95% confidence interval.

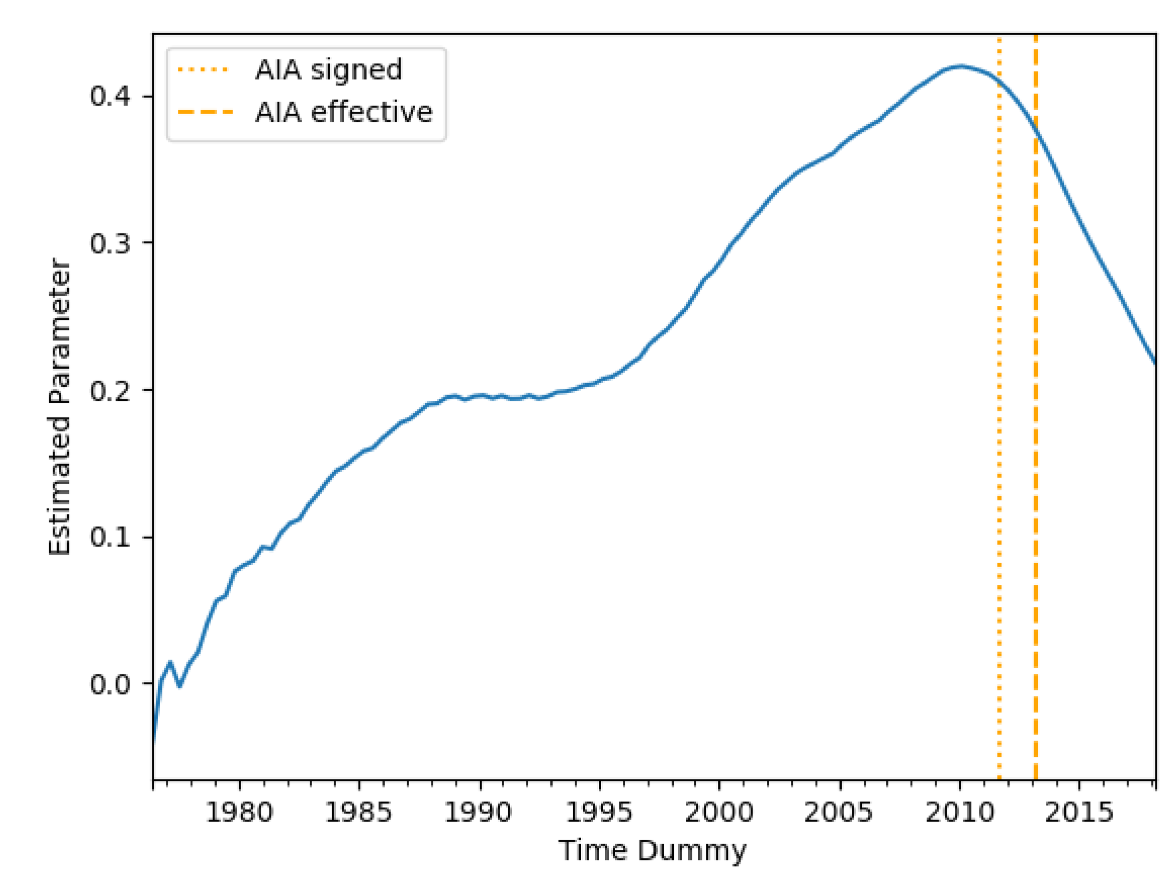

Figure 4: Parameter estimates for time dummy variables from pooled OLS regression with clustered entity coefficients, smoothed.

Indicators

- The number of claims is a good indicator of contribution (Table 1).
- More experienced inventors tend to contribute more.
- U.S. corporations (the withheld dummy) contribute the least; governments contribute most.

Variable	Coeff. (SE)
Log(Num. Claims)	.344 (.007)*
Log(Avg. Inventor Patents)	.536 (.003)*
Assignee: Foreign Co.	.162 (.013)*
Assignee: U.S. Individ.	.866 (.051)*
Assignee: Foreign Individ.	.094 (.070)
Assignee: U.S. Govt.	.799 (.031)
Assignee: Foreign Govt.	.867 (.076)*
Adj. R-Squared	.724

Table 1: Coefficients from an OLS regression of $\log(TKC)$ on patent features. (*) indicates significance at the 0.1% level. Dummy variables for assignee types and NBER categories were also included.

Conclusions

- A novel total knowledge contribution (TKC) index is used to measure the impact of patents on subsequent inventions.
- Knowledge contribution significantly differs across test sectors, especially in newer industries.
- Government organizations produce patents with higher TKC than corporations.
- The AIA had a negative effect on contribution rates, which have decreased sharply with time.
- Future research:
  - Correlate NSF patent funding and knowledge impact.
  - Investigate inter-sector knowledge flow dynamics.

Resources & Acknowledgements

- Scraper, algorithms, and analysis tool: `rbsteed.com/datamaster`.
- Patent data were scraped from the USPTO PatentsView API (`uspto.gov`).
- This work was supported by the GWU Data-MASTER program (NSF Grant 1406984).
- Professor Rahul Simha provided research mentorship.