Open vs. closed innovation: using online network data to measure innovation

Benjamin Snow and Oliver Bott

14 November 2014

Contents

1 Introduction

Policy makers worldwide have a profound interest in innovation for its significance for economic development and prosperity. Taylor (2004) views innovation as "the driving force behind modern economic growth, relative industrial power, and competitive advantage" (p.222). Numerous studies, for example the Innovation Union Scoreboard[1] and the OECD Science, Technology and Innovation Scoreboard[2], have attempted to measure and compare innovation performance on the national level. However, until today most examinations of innovation have put their analytical emphasis on national level patent data, relying on this form of registering of proprietary data as a means of measuring innovation. However, the use of pantent data has widely been criticized for its limited view of innovation, as patents are only filed as a means of protecting an idea for its exclusive commercial use. Many forms of innovation, with the key example being the opensource development of software, do not require or attempt government protection through patents, but would still in a digitalized society be considered as important forms of innovation.

Thus the academic literature until this point leaves largely unexplored other, more open measures of innovation disregarded. The relatively recent emergence of network-based research systems offer new and potentially more instructive metrics by which to measure innovation, compared to protected closed knowledge. Since various scholars called for the continuous improvement of innovation measurement (see for example Freeman and Soete 2009), this work seeks to go beyond the widespread use of patent data to contribute to the refinement of innovation indicators, and the field as a whole.

[1]For the latest edition see [http://ec.europa.eu/enterprise/policies/innovation/policy/innovation-scoreboard/index_en.htm].
[2]For the latest edition see [http://www.oecd.org/sti/scoreboard.htm].

2 State of the Field

2.1 Defining Innovation

Using a rather grand view in his understanding of innovation, Schumpeter (1942) sees innovation as "a process of industrial mutation, that incessantly revolutionizes the economic structure from within". In a more understated characterization, Smith (2005) defines innovation as "the creation of something qualitatively new, via processes of learning and knowledge building. It involves changing competences and capabilities, and producing qualitatively new performance outcomes" (Smith 2005, 149). While it is widely accepted that innovation can take many forms, e.g. product, process, marketing and proccess innovations, Frankelius (2009), in his extensive literature review of innovation studies, criticizes the widespread underlying assumption that innovation is limited to technological innovation. While accepting Frankelius (2009)'s critique of innovation as taking place outside of the technological realm, for the purpose of this study technological innovation, and specifically software innovation, will be the primary focus following relatively closely to Smith (2005)'s definition.

2.2 Measuring Innovation

The frequent technological focus when studying innovation can partly be explained by the difficulties associated with innovation's measurement. Smith (2005) notes the measurement challenge, as innovation is by definition a novelty and thus commensurability is a demanding task. For these reasons innovation has traditionally though controversially been measured by looking either at its inputs, outputs, our throughputs. Attempting to measure innovation by inputs often focuses on resources, such as personnel and equipment allocated to research and development (R&D), which Freeman and Soete (2009) notes is often overestimating innovation in R&D by including the routine with the novel. Put another way, the use of research and development funding to assess innovation assumes innovation to takes place linearly with enough resources, as if the doubling of the number of Austrian patent office workers would have somehow resulted in two Albert Einstein's coming from their ranks, rather that just one. Freeman and Soete (2009) compares this to output oriented measures, which are often based on what he concludes are the already inadequate measures such as GDP. As Freeman and Soete (2009) suggests, GDP is an often cited imprecise statistical measure. However, building off of the first example, the use of output oriented measures the assumed result of innovation, economic growth, and not only assumes that the growth was based upon innovation, but seems to similarly assume that innovation creates value in a linear manner. A simple check of the wealth earned from various patents suggests this is not the case. Thus both input measures, such as R&D funding, and output measures, such as GDP, both either to not directly measure innovation, or do so in a manner so broad as to be unhelpful.

An indicator most often found in innovation research is patent data (see Taylor 2004). A patent is a "temporary legal monopoly granted by the government to an inventor for the commerical use of the invention [Taylor (2004), 229). A patent constitutes a property right awarded when an invention is shown to be non-trivial, useful, and novel (Taylor 2004, 230). Patents were first used to measure demand-side determinants of innovation, and have been used in the analysis of innovation activity for over three decades [Taylor (2004), 230). Taylor (2004) uses patent data taken from 1963 to 1999 in six different industries and their future citation levels and uses Ordinary Least Squares (OLS) model to test what he terms the 'industry-innovation assumption'. The use of citations with patent data suggests a more nuanced examination of innovation using patent data that R&D fuding or GDP, as by using relative citation levels Taylor (2004) was able to weight the relative importance of a patent.

2.3 Limitations of Patent Data

Despite the usefulness of patent data, Taylor (2004) finds several limitations. In addition to the 'classification problem' related to assigning a specific industries to patents, patents vary widely in significance, both technically and economically (Taylor 2004, 231). Most significantly for the purpose on this study, Taylor (2004) as well as Pakes and Griliches (1980) find that "patents are a flawed measure particularly since not all new innovations are patented and since patents differ greatly in their economic impact" (Taylor (2004), 378). Taylor (2004) to some degree is able to take into account the relative importance of different patents by taking into account their future citation level, a relatively good proxy for importance. Still he is less able to tackle the problem of new innovations which are not patented, with patent levels subsequently underrepresenting innovation. Thus, while for some considerable time patents have been considered to be the most effective proxy with which to measure innovation, they themselves ascribe to the notion that there is room for improvement in the study of innovation. This study stands as an attempt to further this field, to attempt to diliniate innovation which would not appear in patent data, but is instead based upon open network data.

2.4 Using Network Data as Innovation Indicator

Current developments in research indicate that "characteristics that were important last century may well no longer be so relevant today and indeed may even be positively misleading" (Freeman and Soete 2009, 3). A shift away from the belief that innovation only occurs in professional R&D labs has occured, a change towards what Freeman and Soete (2009) calls "research without frontiers" (p.13). Even though networks and research collaborations become increasingly important, there have been relatively few studies focusing on network data (see Breschi and Malerba 2005). Even where research networks have been analyzed, the focus is too often on economically useful knowledge (see Acs, Anselin, and Varga 2002). Other studies focusing on research networks focus on other protected collaborative networks (see Ponds, Van Oort, and Frenken 2010). In an exception to this standard, Senghore et al. (2014) attempt to answer whether social network statistics act as indicators of innovation performance within a network, and which statistics could predict innovation performance. Using Gnyawali and Srivastava (2013)'s use models on cluster and network effects to analyze multipartite social networks at mass collaboration events, gathering their data from NASA's International Space Apps Challenge. They use graph theory models constructed from affiliation networks finding (preliminarily) that distributions likely correlate to key aspects (Senghore et al. 2014).

Since Freeman and Soete (2009), among others, calls for the continuous improvement of innovation measurement, this work seeks to contribute to the refinement of innovation indicators. The purpose of this study is to explore the conceptual and statistical viability of a new metric by which we can measure innovation.

2.5 Research Question

In light of the above mentioned state of innovation research we plan to examine the following research question: *To what extent can open innovation network data add to the measurement of innovation performance?*

Exploiting technological advances related to the increasing use of the internet and open research platforms like GitHub, we plan to explore whether open knowledge networks can help refine currently limited innovation performance measurements.

3 Methodology

3.1 Data Gathering

To examine open network data against patent data, this study relies on two key data sources and uses the statistical tool *R* (R Core Team 2014) for the data analysis. We take city-level data of 135 cities overall, ranging from sixteen different countries.

API network data

The first data set is obtained by using the Application Programming Interface (API) data for open networks. To examine open innovation, data is obtained from the the git repository web-based hosting service GitHub[3]. Its use of source code management makes it a commonly used software development collaboration tool. Since most of the repositories are openly accessible one can use API tools to track the popularity of contributors through a process called following. The *R* (R Core Team 2014) packages *httr* (Wickham 2014a), *dplyr* (Wickham and Francois 2014) and *rjson* (Couture-Beil 2014) allow for compiling data on the follower counts and locations associated with different users and online reponsitories.

In our analylsis, we compile follower data in various categories ranging from users with 1-9, 10-19 up to 90-99 and over 100 followers (per 10,000 population). We choose GitHub users with more than 20 followers as our main indicator of high collaboration and innovative activity.

[3]Online accessilbe on [https://github.com/].

Closed innovation OECD patent data

For closed innovation we use city-level patent data, taken from the Organization for Economic Co-operation and Development[4]. Patent Cooperation Treaty (PCT) patent data are used to track internationally patented inventions. The *R* (R Core Team 2014) package *rsdmx* (Blondel 2014) is necessary for obtaining the OECD dataset. As we are interested in patent data, we work with data indicating the PCT patent applications per 10,000 inhabitants. From the same database, we also use GDP per capita data and environmental data, as other variables thought potentially relevant in explaining differences in innovation.

To allow for reproducibility of this research, the code for gathering and cleaning the data is stored in a separate .R file and can be accessed here. Packages to clean, analyze and visualize the data include *R-car* (Fox and Weisberg 2011), *R-ggmap* (Kahle and Wickham 2013), *ggplot2* (Wickham and Chang 2014), *maps* (Brownrigg 2014), *maptools* (Bivand and Lewin-Koh 2014), *Rcpp* (Eddelbuettel and Francois 2014), *rCharts* (Vaidyanathan 2013), *RCurl* (Temple Lang 2014), *repmis* (Gandrud 2014), *reshape2* (Wickham 2014b) and *stargazer* (Hlavac 2014).

3.2 Data Sources

As can be seen in Table 1, the analysis is based on cross-sectional data with varying time frames. There are several limitations to the data used in this study. First, the time discrepency between different aspects of the data used, which range from 2005 to 2014, reflect an obvious data comparability constraint. The GitHub user data is taken from December 2014. Secondly, in this analysis for data availability and access reasons, there are some prominent innovation hubs excluded, including San Fransisco and New York. Any found significance will need to take this into account. Last, and perhaps most significantly, when comparing the two measures of innovation, it should be noted that patent data reflects innovation across all types of sectors, whereas Github data mainly reflects innovation within the software technology domain.

`Table 1: Data sources and explanations.`

[4] Online accessible on [http://stats.oecd.org].

Variables	Explanation	Year	Source
Patents	PCT patents per 10,000 population	2008	OECD
GDP	GDP per capita	2008	OECD
Population	Total urban population	2008	OECD
Greenspace	Green area per capita in square metres	2008	OECD
Employment	Employment of metropolitan area as % of national value	2008	OECD
Pollution	Annual average of pop exposure to air pollution PM2,5 in µg/m^3	2005	OECD
Users	GitHub users per 10,000 population	2014	GitHub

3.3 Data Selection

Several potential explanatory variables are collected besides the patent and GitHub data. These variables were selected as they were thought to potentially show cause for why innovation, be it open or closed, occurs in a certain city, but needed to be variables that would not introduce endogeneity to the model.

Greenspace: The Greenspace indicator is deemed potentially relevant in that with a choice of city to innovate in (assuming some level of geographic labor flexibility) there might be a recreational value necessary for attracting talent. Put another way, green cities could attract innovators.

Pollution: The Pollution indicator is taken both as a broad proxy for industrialization (leaving aside a discussion of to what degree pollution is from industry vs cars), that it seemed worth exploring whether a certain level of pollution discouraged talent attraction of innovators on the city-level.

Employment: The Employment indicator is taken largely as an indication of that city´s significance within its national context. Understanding whether a city would likely be viewed as the most prominent or significant, and whether this effects innovation, or whether innovation takes place in smaller provincial cities, is worther understanding. Additionally, seeing if the type of innovation (open vs. closed) depends on the significance of the city is viewed as relevant.

GDP: A GDP indicator explores whether the size of the economy, or wealth generally, is related to innovation on the city-level, and if it is indicator of one type of innovation over another.

Population: A Population variable explores whether there is a necessary city size threshold which corresponds to innovation, and also is taken for controlling for across cities, to find patent data or GitHub data per a number of people in a city.

Follower cutoff point

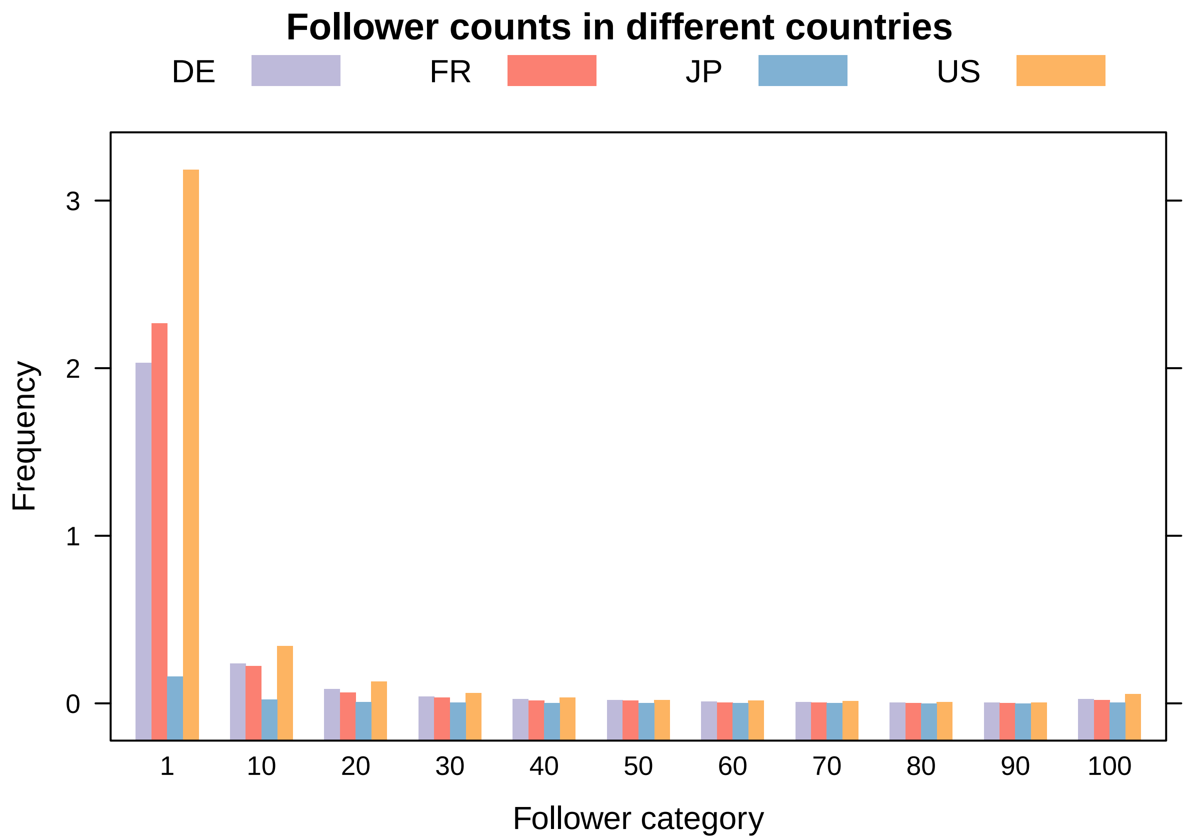

Figure 1:

4 Analysis

4.1 Descriptive Statistics

The summary statistics in Table 2 show wide ranging distributions of the observations in the data set. Since the data cleaning eliminated all values equal to or lower than zero, a log transformation seems to be a strategy that could strengthen the analysis. The *car* package (Fox and Weisberg 2011) is used to examine the relationship, distribution, and normality of all variables included in the model, to understand which regression model would be most appropriate. The distribution of many variables are highly skewed (see Figure 1). All of the GitHub based variables ‘nofollowing’, ‘medfollowing’, and ‘hifollowing’ have significant right skews, as do nearly all of the observed variables, excluding Pollution, GDP and Patents, which come closer to a normal distribution. It seems as if already in the scatterplot a slight correlation between Patents and Followers can be observed. To normalize for the skewed distributions, the log of the variables is deemed necessary to increase the explanatory power of our inferential statistics.

Table 2: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Patents	132	1.69	1.39	0.06	6.86
GDP	132	36,308.69	8,575.53	5,133.12	61,804.14
Population	132	2,062,713.00	3,565,884.00	444,432	34,482,742
Greenspace	132	594.05	936.45	1.13	5,081.25
Employment	132	4.20	7.68	0.15	39.39
Pollution	132	16.21	5.09	5.85	31.44
Users with >0 Followers	132	2.82	3.02	0.00	17.67
1-9 Followers	132	2.32	2.43	0.00	14.34
10-19 Followers	132	0.25	0.29	0.00	1.57
20-29 Followers	132	0.09	0.12	0.00	0.64
30-39 Followers	132	0.05	0.06	0.00	0.37
40-49 Followers	132	0.03	0.04	0.00	0.24
50-59 Followers	132	0.02	0.02	0.00	0.12
60-69 Followers	132	0.01	0.02	0.00	0.11
70-79 Followers	132	0.01	0.01	0.00	0.07
80-89 Followers	132	0.01	0.01	0.00	0.06
90-99 Followers	132	0.004	0.01	0.00	0.05
>100 Followers	132	0.03	0.05	0.00	0.32
Users with >20 Followers	132	0.24	0.33	0.00	1.80

The residual plot between patents and users with high follower numbers (see Figure 2) depicts a relatively random pattern, which indicates that a linear regression model provides a decent fit to the inferential statistics of the data set.

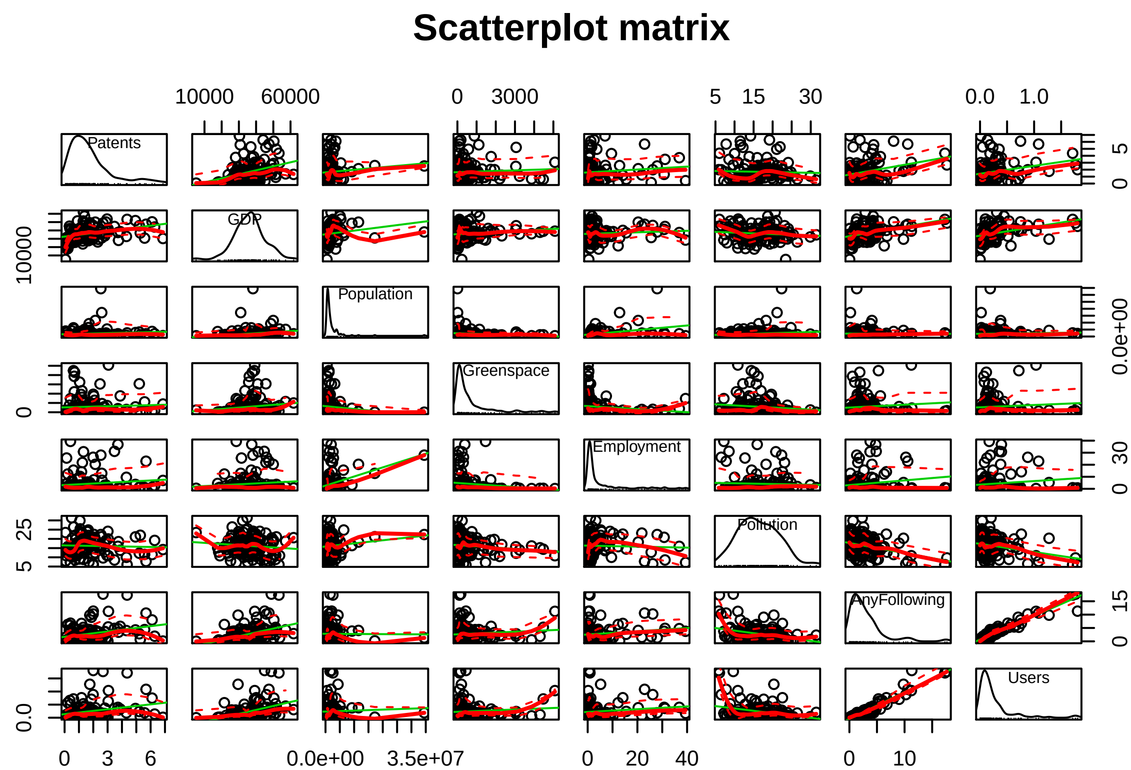
Scatterplot matrix
Patents
GDP
Population
Greenspace
Employment
Pollution
AnyFollowing
Users
10000
60000
0
3000
5
15
30
0.0
1.0
5
0
10000
0.0e+00
0
30
0
25
5
15
0
0.0
0
3
6
0.0e+00
3.5e+07
0
20
40
0
10

Figure 2:

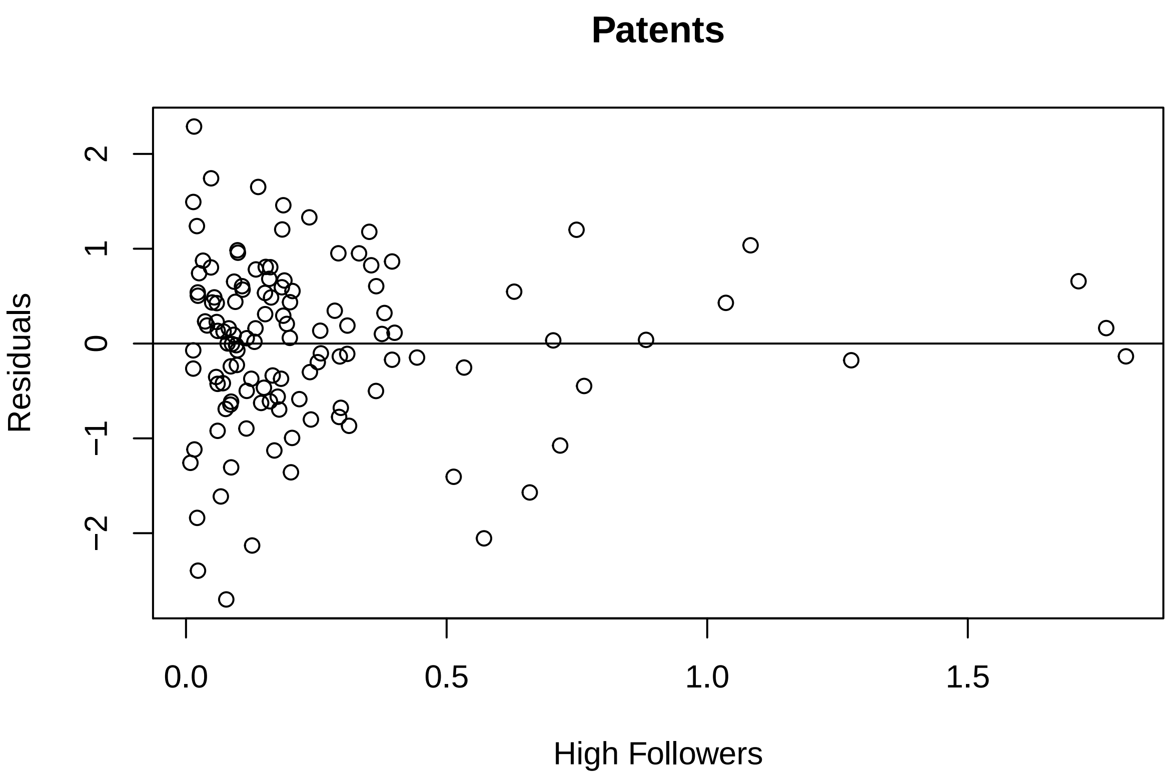

Figure 3:

4.2 Inferential Statistics

As a relationship between patent and follower data is observed, further inferential statistical analysis attempts to find the common predictor or cause of innovation in both patent and open data. The second regression model hence includes the open innovation indicator now as the dependent variable F and is expressed below using similar notation and logic as stated above:

$$logF_l = \beta_0 + \beta_1 logP_l + \beta_2 logGDP_l + \beta_3 logPop_l + \beta_4 logG_l + \beta_4 logE_l + \beta_4 logPol_l + \epsilon_l$$

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu

% Date and time: Do, Dez 11, 2014 - 10:51:48

Table 3: Regression Estimates of GitHub Collaborative Activity

	Dependent variable:		
	log(AnyFollowing)		
	(1)	(2)	(3)
Users	0.37*** (0.11)	0.03 (0.12)	0.16 (0.12)
GDP		1.66*** (0.44)	1.04** (0.43)
Population		−0.02 (0.14)	−0.10 (0.19)
Greenspace		0.18*** (0.06)	0.02 (0.06)
Employment		0.14* (0.08)	0.30* (0.17)
Pollution		−0.95*** (0.28)	−0.63** (0.28)
US			0.88 (0.57)
DE			0.36 (0.38)
FR			0.59 (0.36)
JP			−1.93*** (0.49)
(Intercept)	0.44*** (0.11)	−15.10*** (4.40)	−7.95 (5.24)
Observations	129	129	129
R^2	0.08	0.34	0.52
Adjusted R^2	0.07	0.31	0.47
Residual Std. Error	1.22 (df = 127)	1.05 (df = 122)	0.92 (df = 118)
F Statistic	10.40*** (df = 1; 127)	10.58*** (df = 6; 122)	12.53*** (df = 10; 118)
Note:			*p<0.1; **p<0.05; ***p<0.01

As can be seen in Table 4, again Patents are strongly positively correlated with Follower numbers (at a significance level of $p<0.01$). Both Pollution and Employment are negatively correlated (at a significance level of $p<0.05$). The adjusted R squared value indicates that about 36% of the variation in follower numbers in our sample is explained through the model.

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu

% Date and time: Do, Dez 11, 2014 - 10:51:48

Table 4: Regression Estimates of Patent Activity

	Dependent variable:		
	log(Patents)		
	(1)	(2)	(3)
Over 20 Followers	0.25*** (0.07)	0.08 (0.08)	0.08 (0.08)
GDP		1.90*** (0.40)	1.90*** (0.40)
Population		−0.11 (0.11)	−0.11 (0.11)
Greenspace		0.05 (0.05)	0.05 (0.05)
Employment		0.11* (0.06)	0.11* (0.06)
Pollution		0.45** (0.21)	0.45** (0.21)
(Intercept)	0.68*** (0.16)	−19.58*** (3.82)	−19.58*** (3.82)
Observations	123	123	123
R^2	0.09	0.31	0.31
Adjusted R^2	0.08	0.27	0.27
Residual Std. Error	0.87 (df = 121)	0.78 (df = 116)	0.78 (df = 116)
F Statistic	12.30*** (df = 1; 121)	8.61*** (df = 6; 116)	8.61*** (df = 6; 116)
Note:			*p<0.1; **p<0.05; ***p<0.01

This study plans to use an ordinary least squares (OLS) model to examine the relationship between patent and highly followed open data sources. The model for the regression analysis can be viewed as:

$$logP_i = \beta_0 + \beta_1 logF_i + \beta_2 logGDP_i + \beta_3 logPop_i + \beta_4 logG_i + \beta_5 logE_i + \beta_6 logPol_i + \epsilon_i$$

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu

% Date and time: Do, Dez 11, 2014 - 10:51:49

Here P is the patent intensity expected in a given city i. As seen in the regression output Table 3, a positive relationship between patent data and GitHub data is observed (at a significance level of p<0.01), though most significantly between patent data and those with high numbers of followers on GitHub. In the full model specification, a 1 percent increase in GitHub users with more than 25 followers (per 10,000 population) corresponds with a 1.76 percent increase in PCT patents (per 10,000 population). Additionally, Employment seems also to be positively correlated with patents, supporting the initial hypothesis that the significance of a city in a national context is strongly related to patent activities. Pollution is negatively correlated with

Table 5: Regression Estimates of GitHub Collaborative Activity

	Dependent variable:		
	log(Users)		
	(1)	(2)	(3)
Patents	0.37^{***} (0.11)	0.11 (0.11)	0.11 (0.13)
GDP		2.10^{***} (0.49)	1.54^{***} (0.58)
Population		−0.08 (0.14)	−0.22 (0.20)
Greenspace		0.05 (0.06)	−0.01 (0.06)
Employment		0.09 (0.07)	0.34^{*} (0.17)
Pollution		-0.74^{***} (0.25)	-0.60^{**} (0.28)
US			1.01^{*} (0.58)
DE			0.56 (0.39)
FR			0.65^{*} (0.38)
JP			−0.48 (0.57)
(Intercept)	-1.99^{***} (0.10)	-21.10^{***} (4.71)	-13.99^{**} (6.29)
Observations	123	123	123
R^2	0.09	0.32	0.36
Adjusted R^2	0.08	0.28	0.30
Residual Std. Error	1.06 (df = 121)	0.93 (df = 116)	0.92 (df = 112)
F Statistic	12.30^{***} (df = 1; 121)	9.01^{***} (df = 6; 116)	6.31^{***} (df = 10; 112)
Note:			*p<0.1; **p<0.05; ***p<0.01

patent data at a high significance level, while the variables GDP, Population and Greenspace do not seem to have a significant effect on patent activity.

The findings endorse the implicit hypothesis of the study that open data sources seem to show innovation in a similar but perhaps distinct manner to patent data and could hence enrich the measurement of innovation. These found relationships offer a glimps into the ‘throughput’ of open innovation (in the form of collaboration) rather than the ‘output’ which patent data reflects (in the form of commercialization of knowledge). While significant relationships are found with innovation, other unaccounted for variables can be expected to contribute but are not currently accounted for in the model.

5 Discussion

6 Outlook

From this analysis it becomes apparent that the introduction of various dummy controls could help explain the spurious relationsbetween patent and network data. This could also help to find the more fundamental factors influencing innovation activity. Hence, to better answer the stated research question, it seems sensible to control for English speaking countries, as one would suspect the spread of GitHub to be greatest there. Also, one could introduce dummy controls for the overall economic development of the country, assuming that software development is clustered in these locations. In addition, including a map visualization with information on location of the cities in the sample could improve this work.

References

Acs, Zoltan J, Luc Anselin, and Attila Varga. 2002. "Patents and Innovation Counts as Measures of Regional Production of New Knowledge." *Research Policy* 31 (7). Elsevier: 1069–85.

Bivand, Roger, and Nicholas Lewin-Koh. 2014. *Maptools: Tools for Reading and Handling Spatial Objects*. http://CRAN.R-project.org/package=maptools.

Blondel, Emmanuel. 2014. *Rsdmx: Tools for Reading SDMX Data and Metadata*. http://CRAN.R-project.org/package=rsdmx.

Breschi, Stefano, and Franco Malerba. 2005. *Clusters, Networks and Innovation*. Oxford University Press.

Brownrigg, Ray. 2014. *Maps: Draw Geographical Maps*. http://CRAN.R-project.org/package=maps.

Couture-Beil, Alex. 2014. *Rjson: JSON for R*. http://CRAN.R-project.org/package=rjson.

Eddelbuettel, Dirk, and Romain Francois. 2014. *Rcpp: Seamless R and C++ Integration*. http://CRAN.R-project.org/package=Rcpp.

Fox, John, and Sanford Weisberg. 2011. *An R Companion to Applied Regression*. Second. Thousand Oaks CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Frankelius, Per. 2009. "Questioning Two Myths in Innovation Literature." *The Journal of High Technology Management Research* 20 (1). Elsevier: 40–51.

Freeman, Christopher, and Luc Soete. 2009. "Developing Science, Technology and Innovation Indicators: What We Can Learn from the Past." *Research Policy* 38 (4). Elsevier: 583–89.

Gandrud, Christopher. 2014. *Repmis: A Collection of Miscellaneous Tools for Reproducible Research with R*. http://CRAN.R-project.org/package=repmis.

Gnyawali, Devi, and Manish Srivastava. 2013. "Complementary Effects of Clusters and Networks on Firm Innovation: A Conceptual Model." *Journal of Engineering Management*, no. 30: 1–20.

Hlavac, Marek. 2014. *Stargazer: LaTeX/HTML Code and ASCII Text for Well-Formatted Regression and Summary Statistics Tables*. http://CRAN.R-project.org/package=stargazer.

Kahle, David, and Hadley Wickham. 2013. *Ggmap: A Package for Spatial Visualization with Google Maps and OpenStreetMap*. http://CRAN.R-project.org/package=ggmap.

Pakes, Ariel, and Zvi Griliches. 1980. "Patents and R&D at the Firm Level: A First Report." *Economics Letters* 5 (4). Elsevier: 377–81.

Ponds, Roderik, Frank Van Oort, and Koen Frenken. 2010. "Innovation, Spillovers and University–industry Collaboration: An Extended Knowledge Production Function Approach." *Journal of Economic Geography* 10 (2). Oxford Univ Press: 231–55.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Schumpeter, Joseph. 1942. "Creative Destruction." *Capitalism, Socialism and Democracy*.

Senghore, Fatima, Enrique Campos-Nanez, Pavel Fomin, and James S Wasek. 2014. "Using Social Network Analysis to Investigate the Potential of Innovation Networks: Lessons Learned from NASA's International Space Apps Challenge." *Procedia Computer Science* 28. Elsevier: 380–88.

Smith, K. H. 2005. "Measuring Innovation." PhD thesis, Oxford University Press.

Taylor, M. Z. 2004. "Empirical Evidence Against Varieties of Capitalism's Theory of Technological Innovation." *International Organization* 58 (03). Cambridge Univ Press: 601–31.

Temple Lang, Duncan. 2014. *RCurl: General Network (HTTP/FTP/.) Client Interface for R*. http://CRAN.R-project.org/package=RCurl.

Vaidyanathan, Ramnath. 2013. *RCharts: Interactive Charts Using Javascript Visualization Libraries*.

Wickham, Hadley. 2014a. *Httr: Tools for Working with URLs and HTTP*. http://CRAN.R-project.org/package=httr.

———. 2014b. *Reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. http://CRAN.R-project.org/package=reshape2.

Wickham, Hadley, and Winston Chang. 2014. *Ggplot2: An Implementation of the Grammar of Graphics*. http://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, and Romain Francois. 2014. *Dplyr: A Grammar of Data Manipulation*. http://CRAN.R-project.org/package=dplyr.