



Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

Technological interdependencies predict innovation dynamics

Anton Pichler, François Lafond and J. Doyne Farmer

March 2020

INET Oxford Working Paper No. 2020-04

Complexity Economics



Technological interdependencies predict innovation dynamics

Anton Pichler^{a,b,c}, François Lafond^{a,b}, J. Doyne Farmer^{a,b,d}

^aInstitute for New Economic Thinking at the Oxford Martin School, University of Oxford, Manor Road, OX1 3UQ, UK

^bMathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

^cComplexity Science Hub Vienna, Josefstädter Straße 39, A-1080, Austria

^dSanta Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Abstract

We propose a simple model where the innovation rate of a technological domain depends on the innovation rate of the technological domains it relies on. Using data on US patents from 1836 to 2017, we make out-of-sample predictions and find that the predictability of innovation rates can be boosted substantially when network effects are taken into account. In the case where a technology's neighborhood future innovation rates are known, the average predictability gain is 28% compared to simpler time series model which do not incorporate network effects. Even when nothing is known about the future, we find positive average predictability gains of 20%. The results have important policy implications, suggesting that the effective support of a given technology must take into account the technological ecosystem surrounding the targeted technology.

Keywords: innovation, technology, network, forecasting, patents, spatial econometrics

1. Introduction

Technological evolution is often described as a recursive process whereby the recombination of existing components leads to new or improved technological components (Schumpeter, 1939; Usher, 1954; Kauffman, 1993; Fleming, 2001; Arthur, 2009; McNeerney et al., 2011; Tria et al., 2014; Youn et al., 2015; Fink and Reeves, 2019). A simple hypothesis, therefore, is that technological domains that tend to recombine elements from fast-growing technological domains should themselves grow faster. In other words, a technology will tend to progress faster if the technologies it relies on are themselves making fast progress. While these ideas are well established, very little has been done to establish empirically that technological interdependencies help predict future innovation dynamics. Being able to demonstrate this relationship would be very helpful, as it would allow us to support key technologies and overall technological progress by designing and supporting technological ecosystems.

In this paper, we establish that knowing the technological ecosystem helps predict the dynamics of future innovation. We use a simple model where the innovation rates of a technological domain depends on efforts within the domain, but also on the stock of knowledge in the domain and in supporting domains. We test the model on the record of United States Patent Office (USPTO) patents from 1836 to 2017, which includes more than 10M patents, 40M classifications in ~ 650 technological domains, and 90M cita-

tions. As predicted by the model, we find that that the growth rates of the number of patents in a technological domain depends strongly on the growth rates of its knowledge sources. Given the volatile nature of growth rates, this strong relationship is remarkable. We use this insight for making out-of-sample predictions of patenting activities and find that integrating network effects can improve predictions substantially compared to independent time series models. The results have important policy implications, suggesting that research policy targeted at fostering innovation in a technological domain has to take its surrounding technological ecosystem into account.

Several studies have put forward the idea of network-dependent innovation dynamics. For instance, Cowan and Jonard (2003) and König et al. (2011) have proposed models of innovation arising from the recombination of knowledge in R&D partnership networks.

Acemoglu et al. (2016) find that upstream patenting levels are highly correlated with downstream patenting levels. Taalbi (2018) finds broadly similar results using innovation counts and a network constructed to reflect which industry uses innovation from another industry. Patenting levels are persistent, and so are automatically very predictable. In contrast, we focus on changes in patenting levels, which are not persistent and are much more difficult to predict.

Our work on innovation dynamics is also related to evolutionary models such as Farmer et al. (1986), Bagley et al. (1992) and Jain and Krishna (2001). In these models, the exponential growth trajectories of nodes arise due to the existence of autocatalytic sets, i.e. a subset of the network where nodes have at least one positive incoming link from

*Corresponding author

Email address: anton.pichler@maths.ox.ac.uk (Anton Pichler)

a node of the same subgraph, leading to sustainable self-reinforcing dynamics. Evidence for the presence of autocatalytic sets in technology systems was recently provided by Napolitano et al. (2018) who find that the autocatalytic structure of the patent network has grown in time and that patent classes belonging to the autocatalytic set show higher levels of innovation activity.

Our contribution to the literature can be summarized as follows. First, we show strong empirical evidence of coupled innovation growth in the presence of network linkages and demonstrate that the observed effects are far from what would be expected by chance (Section 2). Second, we propose a simple theoretical model of network-dependent knowledge production which is able to explain the observed empirical pattern. We then show how the model can be estimated from empirical data, and quantify how network effects in innovation dynamics have become more important over the last 70 years (Section 3). Finally, we validate the model by predicting future patenting levels (Section 4). We find that independent time series models can be substantially improved in case network information is available.

2. Empirical evidence

2.1. Data and network construction

We use data on the whole universe of granted United States patents starting in 1836 up to 2017, where the year refers to the publishing date of the patent. The dataset covers 9.83 million patents and over 91 million citations. Each patent is categorized by the patent office into one of several Cooperative Patent Classification (CPC) codes. We regard each 4-digit CPC code as a distinct technological domain¹. We construct a directed network C_t where an element $C_{ij,t}$ is the sum of all citations from technology j at time t to technology i . In Appendix A we show how this matrix can be derived from patent citations by straightforward algebra². In our analysis, we use row-normalized matrices $\{W_t : t = 1947, \dots, 2017\}$ where each element $W_{ij,t}$ is defined as

$$W_{ij,t} := \frac{C_{ij,t}}{\sum_{j=1}^N C_{ij,t}}. \quad (1)$$

Thus, the row $\{W_{ij,t} : j = 1 \dots N\}$ can be interpreted as technology i 's dependence in its patenting activity at time t on all other technologies: each entry is the share of citations from technology i at time t to another technological domain j . This is a temporal network with yearly snapshots

¹ To check robustness of our results with respect to alternative classification schemes and aggregation levels, we redo the entire analysis presented in the main text on the level of CPC 3-digits, International Patent Classification (IPC) 3- and 4-digits and report the results in the supplementary information.

² The analysis shown here does not rule out alternative technological distance measures. Instead of using technology citation networks, other networks such as co-classification and co-citation networks could be used.

(Masuda and Lambiotte, 2016). We do not construct networks for years before 1947, because earlier citations made by patents are not well documented. But citations to earlier patents are well reported. For example, a citation from a patent granted after 1947 to a patent of the 19th century is included in the network³.

2.2. Preliminary evidence

We now investigate whether technological domains connected to fast-growth technological domains also grow faster. In other words, we study the assortativity of the temporal network with respect to patenting growth rates. For each technology, we compute its neighborhood patenting growth rate as the average nearest neighbor growth rate (ANNG),

$$\tilde{g}_{i,t} = \sum_{j=1}^N \frac{C_{ij,t}(1 - \delta_{ij})}{\sum_{k=1}^N C_{ik,t}(1 - \delta_{ik})} g_{j,t}, \quad (2)$$

where δ_{ij} denotes the Kronecker delta (the matrix diagonal is excluded in the summation so that the ANNG measures neighbor growth rates only). We then test if there is a positive correlation between growth rates of patenting activity in technological domains and their ANNG.

Fig. 1a) plots the 10-year growth rates against the 10-year ANNG for the technology network in 2000, revealing a highly significant positive relationship in growth rates from 2000 to 2010 between technologies and their neighborhood. Since the technology network is based on the year 2000, the figure suggests that if a technology draws a significant share of knowledge from another technology in 2000, we would expect them to exhibit similar growth rates over the next ten years. This pronounced positive relationship is striking given that it is based on growth rates, which tend to be much noisier than patenting levels. While one could expect the network to be assortative with respect to degrees and patenting levels, assortativity with respect to growth rates is far less trivial.

Although interesting, we should test if this relationship holds over time and if we really can be sure that the observed relationship is significantly different from what a null model would produce. To benchmark our results, we calculate the correlations between growth rates and ANNG rates for a randomized version of the given technology network. For each year t and a fixed row i in the matrix W_t , we resample all off-diagonal entries without replacements. In the randomized control network the nodes still have the same weighted outgoing links, but now randomly pointing to other nodes (excluding self-loops). We repeat the

³ Alstott et al. (2017) argue that links of technology networks can be driven by domain sizes and patent ages and suggest a network normalization procedure to control for these impinging factors. We have applied this method as explained in detail in Appendix B, but our results are not sensitive with respect to the network normalization.

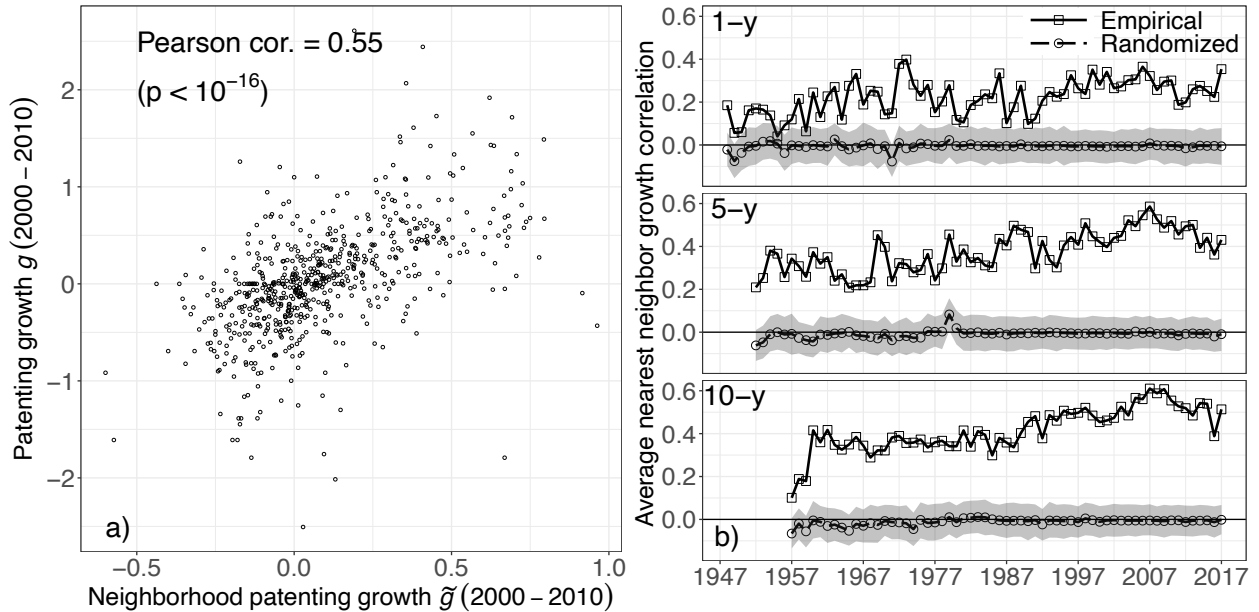


Figure 1: a) Technologies’ 10-year growth rates g for the horizon 2000 to 2010 plotted against average nearest neighbor growth rates \tilde{g} (ANNG). The neighbors’ growth rates are calculated based on the technology network in 2000. The plot thus suggests that technologies experienced a similar growth rate over the next ten year as the technologies they were linked to in 2000. b) Network assortativity over time with respect to 1-year (upper panel), 5-year (center panel) and 10-year (bottom panel) growth rates. There is pronounced correlation between growth rates of nodes and their neighbors in the technology network, while the randomized assortativity measure hovers around zero. The shaded area around the randomized network predictions depict the 0.05-0.95 quantile range of the 1,000 sample realizations. The correlations tend to get stronger in time in the empirical network.

randomization process 1,000 times and report the average correlation between growth rates and ANNG as the corresponding null model. Fig. 1b) plots the Pearson correlation over time between growth rates and ANNG, once computed in the given technology network and once in the randomized control. We report the results for three different growth rates, 1-year (upper panel), 5-year (center panel) and 10-year growth rates (bottom panel). The positive correlation structure between knowledge sources growth and own growth rates over time is far from what one would expect from a network where nodes distribute their out-links randomly over the whole set of potential knowledge sources. Intuitively, one would expect that network effects materialize over the long run instead of showing immediate effects. The figure confirms this hypothesis, but even for 1-year growth rates we find relatively strong and significant positive correlations. Another interesting aspect is that for all three growth rate types, the positive relationship tends to get stronger over time.⁴ Further evidence on the impact of a technology’s neighborhood on its patenting dynamics is discussed in the supplementary

⁴ It should be noted that technology citation networks are by construction correlated with co-classification networks (see Appendix A), and thus, their marginal effects are difficult to disentangle. To test whether there is signal in the citations themselves, we computed the same ANNG measures for USPC main classes. When considering only USPC main classes, there is no co-classification, allowing us to separate these effects. Although slightly noisier, we also find strongly positive ANNG values for the less granular USPC main classes.

information.

3. Network-dependent knowledge creation

3.1. The model

To explain the empirical observations, we now introduce a simple model of network-dependent knowledge creation and discuss its implications for the long-term evolution of technologies. Let us consider N distinct technological domains. The creation of new knowledge in a technological domain requires active research effort, and depends positively on the existing stock of knowledge, from the same domain or from specific other domains.

More precisely, we assume that the creation of new knowledge follows the dynamical system

$$\dot{K}_i(t) = \theta_i R_i(t)^\alpha \prod_{j=1}^N K_j(t)^{\beta W_{ij}}, \quad (3)$$

where θ_i is a technology-specific productivity parameter, $R_i(t)$ the research effort in domain i at time t and $K_i(t)$ is the stock of knowledge in i at t . The technological ecosystem is represented by the weighted adjacency matrix W which is normalized to be row-stochastic and for mathematical convenience assumed to be fixed in time. A technological domain is then a node in the network whose innovation rates depend on its location in the technological ecosystem. If a technology i draws knowledge from the

knowledge stock of node j in the innovation process, the two nodes are connected through a directed edge from i to j . The directed link has the weight W_{ij} denoting the share of technology j 's knowledge in the creation of new knowledge in technology i . θ_i captures the fact that intrinsic characteristics of technologies affect how easy innovation rates can be influenced. $\alpha \geq 0$ denotes the elasticity of knowledge output with respect to research efforts. The elasticity of knowledge output in i (\dot{K}_i) with respect to the knowledge stock in j (K_j) is βW_{ij} , thus β denotes the sum over all domains j of these elasticities (since $\sum_j W_{ij} = 1$).

Eq. (3) also relates to the knowledge production functions used in classical endogenous growth models (Romer (1990), Aghion and Howitt (1990), Grossman and Helpman (1991) and Jones (1995)), but incorporates network effects. It simplifies to the standard Cobb-Douglas knowledge production function in case each technology uses only its own knowledge stock ($W = \mathbb{I}$). Similar equations have been estimated empirically within the ‘‘R&D spillovers’’ literature which estimates the impact of R&D in one entity on outcomes in another entity, such as countries, regions, firms, and sectors (Ertur and Koch, 2007; Hall et al., 2010; Ho et al., 2018).

Research effort is considered to be an exogenous policy variable which we assume to grow at a constant rate $R_i(t) = R_{i,0}e^{\lambda_i t}$. Dividing Eq. (3) by $K_i(t)$, taking logs and the derivative with respect to time, we obtain after rearranging the nonlinear autonomous system

$$\dot{g}_i(t) = \alpha \lambda_i g_i(t) + (\beta W_{ii} - 1)g_i(t)^2 + \beta g_i(t) \sum_{j=1}^N W_{ij} g_j(t) (1 - \delta_{ij}), \quad (4)$$

where $g_i(t) := \dot{K}_i(t)/K_i(t)$ is the growth rate of the knowledge stock. Eq. (4) has been extensively studied without network effects in traditional endogenous growth models and in this case its dynamics are well-understood (Romer (2012), ch.3). When solving the model without network effects, the steady state growth rate (‘‘balanced growth path’’) is $g_i^* = \alpha \lambda_i / (1 - \beta)$ which is globally stable under the standard assumption of $\beta < 1$. Here, the long run-growth rate of knowledge in technology i can only be increased by increasing the long-run growth rates of research efforts in the particular technology.

When network effects are included, a technology’s long-term growth path depends also on the growth rates of other technologies. By setting $\dot{g}_i = 0$ for all i , we find the steady state of the form

$$g_i^* = \frac{\alpha \lambda_i}{1 - \beta W_{ii}} + \frac{\beta}{1 - \beta W_{ii}} \sum_{j=1}^N W_{ij} g_j^* (1 - \delta_{ij}). \quad (5)$$

The steady state growth path for technology i can be understood as a sum of two components. The first part is the idiosyncratic term, which equals the endogenous growth result without network effects. The second component suggests that the long-run growth of a technology depends positively on its neighbors’ growth rates.

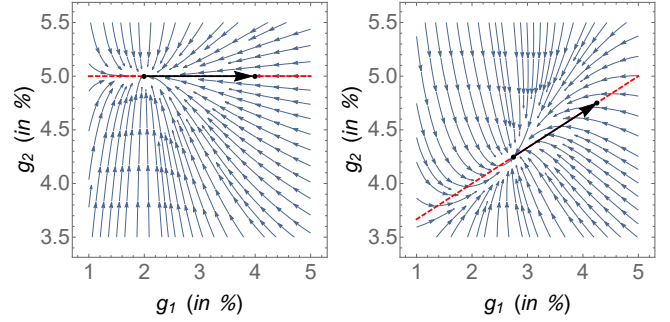


Figure 2: Phase portraits of growth trajectories for a simple example of only two technologies. The horizontal and vertical axes denote the growth rates of technology 1 and technology 2, respectively. In the left panel there are no network effects, $W = \mathbb{I}$. In the right panel a network is present with elements $W_{ij} = 0.5$ for all i, j . The phase diagrams are shown for initial research growth rates $\lambda_1 = 0.02$ and $\lambda_2 = 0.05$. The remaining parameters are set as follows: $\alpha = \beta = 0.5$. With this parametrization all trajectories converge to the steady state $(g_1^*, g_2^*) = (0.02, 0.05)$ in the no-network case and to the steady state $(g_1^*, g_2^*) = (0.0275, 0.0425)$ if network effects are included. In the absence of network effects, the steady state growth rates are simply equal the research effort growth rates. If network effects are included, technology 1 is growing faster in steady state since it draws knowledge from technology 2 which experiences higher research effort growth. In contrast, for technology 2 the steady state growth rate is lower due to the network effects. The black arrows indicate how the steady state of the system moves if research effort growth λ_1 is doubled to $\lambda_1' = 0.04$, but leaving $\lambda_2 = 0.05$ unchanged. In the left panel, this increases only the steady state growth rate for technology 2, but leaves technology 1 unaffected. In the right panel where network effects are included both technologies benefit from the increase in research effort in technology 1. The red dashed line shows the trajectory of the steady state as a function of λ_1 .

To see the difference between the network-dependent model and a simple endogenous growth model version without network effects, let us assume that we could increase the constant growth rate of research effort in a technology i from λ_i to λ_i' . Without network effects, this would simply increase the growth rate of i by $\alpha(\lambda_i' - \lambda_i)/(1 - \beta)$ with no impact on other technologies. Yet if network effects are included, this initial increase of i 's growth rate will also impact neighboring technologies which draw upon the knowledge stock of i , which in turn, will again affect their downstream neighbors and so on. If node i points to any of the affected technologies, it will again experience a change in its growth rate and trigger the process again. We see that the convergence (if any) to the steady state after a shock in the research effort variable is more involved if the innovation rate of a technology depends on other technologies. The phase portraits depicted in Fig. 2 visualize how a change in research efforts impact innovation dynamics in a simple system of only two technologies, one with network effects (right panel) and one without (left panel).

3.2. Calibration

While the dynamical system can have a non-trivial transient, for simplicity we calibrate the model based on the steady state results, using maximum likelihood. To do so,

we reformulate Eq. (5) as the econometric model

$$g_{i,t} = \frac{a_i}{1 - \beta W_{ii,t}} + \frac{\beta}{1 - \beta W_{ii,t}} \sum_{j=1}^N W_{ij,t} g_{j,t} (1 - \delta_{ij}) + \epsilon_{i,t}, \quad (6)$$

where $\epsilon_{i,t} \sim N(0, \sigma^2)$ and a_i is capturing the composite variable $\alpha \lambda_i$.

The specification is related to spatial econometric models using panel data (Elhorst, 2014; Wang and Yu, 2015), but there are also significant differences. To estimate the parameters, we therefore derive an estimator which is outlined in more detail in Appendix C. We allow the technology network to be time-varying, since we observe changing citation networks over time. In the supplementary information we present results of extensive robustness checks, covering alternative model specifications such as using time-fixed networks and spatial autoregressive models. There we also control for further possible explanatory variables, such as technological domain sizes.

The model is derived in terms of knowledge stocks $K_{i,t} = \sum_{\tau=0}^t P_{i,\tau}$, but knowledge stocks grow very smoothly, leaving little variance to exploit for estimating the parameters of interest. In the steady state, however, by definition growth rates are constant over time within each domain. In that case, the (deterministic) growth rates of stocks and flows are the same. Thus, for empirical convenience, we let $g_{i,t} := \ln(P_{i,t}/P_{i,t-\tau})$ be the τ -year patenting growth rate of the technology class i at a time t .

We fit the model to the whole time series up to 2017 and report the estimated parameters in Table 1. Since the results of Section 2 suggest varying magnitudes of network effects for different time lags, we estimate the parameters for 1-year, 5-year and 10-year growth rates (columns one to three, respectively). Unsurprisingly, we find research growth parameters $a_i = \alpha \lambda_i$ which are positive on average and higher for larger time lags. The highly significant network parameter β is large, ranging from 0.85 to 0.94, depending on the growth rate lag. This value is large, because if $\beta > 1$, we would expect exploding dynamics where an increasing knowledge stock leads to ever larger growth rates. β is smaller, but close to one, exemplifying the importance of the existing technology network for future innovation dynamics.

To explore the network impact on innovation growth more systematically, we rewrite the derived result in Eq. (6) into matrix notation,

$$g^* = \alpha L \lambda, \quad (7)$$

where $L := [\mathbb{I} - \beta W]^{-1}$. The matrix representation is useful as it shows how the long-run trajectory of innovation in a given technological domain depends on the research efforts in the entire technological ecosystem. In particular, if research effort is subject to policy, we can use Eq. (7) to study two interesting policy experiments.

First, Eq. (7) allows us to identify the key supporting technologies for each technology. The naïve way to

	Dependent variable: patenting growth		
	(1-y)	(5-y)	(10-y)
average \hat{a}_i	0.01	0.03	0.04
$\hat{\beta}$	0.85*** (0.0059)	0.93*** (0.0060)	0.94*** (0.0076)
$\hat{\sigma}^2$	0.10	0.19	0.29
Log-likelihood	-13,383	-5,464	-3,612
Observations	43,651	8,734	4,348
Note:	***p < 10 ⁻¹⁶		

Table 1: Results from estimating the econometric network model (Eq. 6). The results are shown for 1-, 5- and 10-year growth rates.

foster innovation in a technology is to increase research efforts in the particular technological domain. But Eq. (7) shows that knowledge growth depends also on research in other technologies and that there could even be cases where knowledge spillovers from research in other technologies will be more beneficial than simply devoting additional research efforts in the focal technology’s domain. Knowledge spillovers from technology j to technology i are made explicit when looking at the Jacobian matrix

$$\frac{\partial g_i^*}{\partial \lambda_j} = \alpha L_{ij}. \quad (8)$$

The matrix element L_{ij} therefore informs us on how much we would expect the long-term innovation growth rate of technology i to change as a consequence of a marginal change in technology j . A large row sum of off-diagonal elements $\sum_{j=1}^N L_{ij}(1 - \delta_{ij})$ indicates a large dependence of technology i ’s growth on external research activities. Ordering technologies based on their size in a given column yields a ranking from the most to least important supporting technologies for the focal technology.

As a second policy experiment, we can ask how a sustained change in R&D in a particular sector affects the long-run growth rate of the entire technological ecosystem. This can be relevant for a policy-maker who wants to devote her research investments in an efficient manner. For example, if a choice has to be made in what technology to invest for fostering overall innovation activities, the decision can be supported with Eq. (7). If an economy’s total knowledge is the sum of sectoral knowledge stocks $K_{tot}(t) = \sum_{i=1}^N K_i(t)$, the growth rate of the total knowledge stock can be expressed as $g_{tot}(t) = \sum_{i=1}^N g_i(t) \frac{K_i(t)}{K_{tot}(t)}$. The impact on total knowledge growth as a consequence of a sustained change in research effort growth in a single domain

⁵ When estimating the model based on 1-year growth rates, we find five technologies with larger off-diagonal entries than diagonal elements. Although these technologies are very different, ranging from multipurpose hand tools (B25F) to emergency protective circuit arrangements (H02H), they all rely heavily on the general purpose technology class A61B, diagnosis, surgery and identification, which exhibits the largest downstream spillovers, $\sum_j L_{ij}(1 - \delta_{ij})$, of all technologies.

i is then given by

$$\frac{\partial g_{tot}^*}{\partial \lambda_j} = \alpha \sum_{i=1}^N L_{ij} \frac{K_i}{K} + \alpha \lambda_j \sum_{i=1}^N \frac{\partial(K_i/K)}{\partial \lambda_j} L_{ij}. \quad (9)$$

The impact on overall innovation growth is again a sum of two components. The first part in the sum takes the form of a output multiplier, analogous to the output multiplier in input-output economics. In contrast to traditional input-output economics (Leontief, 1936; Miller and Blain, 2009), however, the output multiplier has to be weighted with the relative domain sizes since the model is in growth rates instead of levels. Increasing research efforts in a domain with a high weighted output multiplier entails large positive effects on overall innovation rates. The second part in the sum accounts for the fact that the weights themselves change due to changes in research effort.

4. Predicting innovation rates

The empirical evidence for network effects on innovation dynamics is strong. In this section we take advantage of this finding to test the predictive power of the estimated network model. We use the model to make out-of-sample forecasts of patenting levels and evaluate their performance by comparing them with predictions based on time series models which do not incorporate network effects explicitly. We test the network model in two separate prediction exercises. In the first prediction exercise we ask whether knowledge on innovation dynamics around a focal technology can help predicting its growth rates. Or to put it differently, conditional on growth rates of i 's neighbors, what is i 's growth rate? We call these the conditional forecasts. To make this more precise, we define the conditional predictor as

$$\hat{g}_{i,t+\tau}^{cond.} := \frac{\hat{a}_i}{1 - \hat{\beta} W_{ii,t}} + \frac{\hat{\beta}}{1 - \hat{\beta} W_{ii,t}} \sum_{j=1}^N W_{ij,t} g_{j,t+\tau} (1 - \delta_{ij}). \quad (10)$$

We also make forecasts where we do not assume knowledge on the neighbors' growth rates, and use only information available up to t . We call these the unconditional forecasts. As explained in detail in Appendix D, predictions in this case are based on the estimator

$$\hat{g}_{t+\tau}^{uncond.} := \left(\sum_{k=0}^{k'} \hat{\beta}^k W_t^k \right) \hat{a} \approx L(\hat{\beta}) \hat{a}, \quad (11)$$

where the approximation is exact in case $k' \rightarrow \infty$. The predictions are essentially agnostic with respect to the used technology network. While our focus lies on knowledge spillovers, the model is easily extendable to alternative technological distance measures, such as co-classification or co-citations networks.

The forecasts of both prediction exercises are compared to the forecasts obtained from standard ARIMA($p,1,q$)

time series models

$$g_{i,t} = \nu_i + \sum_{s=0}^p \phi_{i,s} g_{i,t-s} + \sum_{s=0}^q \psi_{i,s} u_{i,t-s} + u_{i,t}, \quad (12)$$

where $\phi_{i,0} = \psi_{i,0} = 0$. Note that an ARIMA simply reduces to a geometric random walk if $p = q = 0$. As discussed in more depth in the Appendix D, we choose p , q and k' such that the models yield the best forecasting performance in a validation set. The chosen parameters are shown in detail in the supplementary information. Note that predictions from the ARIMA model are the same for both prediction exercises, since here forecasts for a specific domain i are completely independent of other domains.

We estimate all models based on the data between 1947 and 2002 and use the result to predict yearly patenting from 2003 to 2017. To assess the predictive performance of the network models, we calculate the predictability gain for each year and technology as

$$PG_{i,t} = \frac{|P_{i,t} - \hat{P}_{i,t}^{ARIMA}|}{P_{i,t}} - \frac{|P_{i,t} - \hat{P}_{i,t}^{network}|}{P_{i,t}}, \quad (13)$$

where $\hat{P}_{i,t}^{ARIMA}$ denotes the predicted number of patents from the ARIMA model and $\hat{P}_{i,t}^{network}$ the predictions from the network models. Eq. (13) is simply the difference between absolute percentage errors from the ARIMA and the network model predictions.

When taking time averages of Eq. (13) for each time series, we find for 86% of technologies positive mean predictability gains in the conditional forecasts. In the unconditional forecasts, we get positive predictability gains in 63%. The result looks similar when taking time medians instead of averages where we find positive predictability gains for 84% of all conditional forecasts and for 62% of unconditional forecasts. For each technology, we also conduct a one-sided t-test to check if the predictability gains are significantly greater than zero. We find that predictability gains are significantly larger than zero (on the 5% level: note that we have only 15 observations per series) in 69% of all cases for the conditional forecasts and in 41% for unconditional forecasts. On the other hand, only 6% of all time series are found to have significantly negative predictability gains in the conditional forecasting setup and 18% in the unconditional one.

The average predictability gains for each year are shown in Fig. 3, demonstrating the substantial predictive power of network model. In the conditional forecasts, the predictability gain is relatively small for one year growth rates ($\approx 3\%$), but reaches a maximum of roughly 63% in 2009. Note that if a technology's evolution is not influenced by its surrounding technological ecosystem, we would not expect any systematic predictability gain at all. Obviously, it is harder to beat the ARIMA models in the unconditional forecasting scenario where no information on future innovation dynamics of neighboring technologies is available. Nevertheless, the network model performs significantly

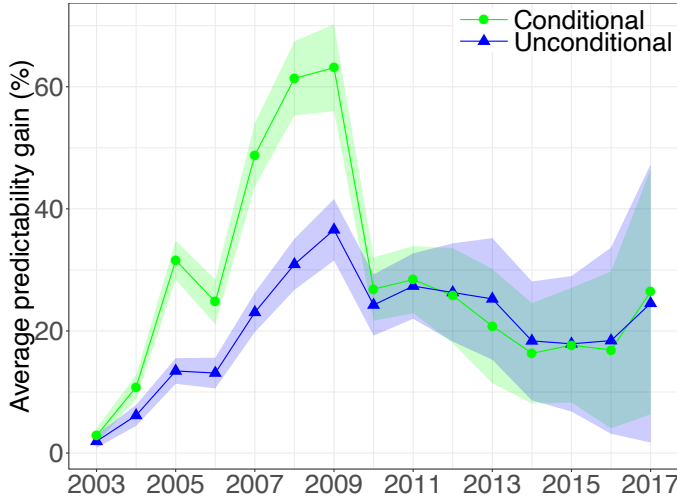


Figure 3: Average predictability gain in %, $\sum_i PG_{it}/N$, over ARIMA time series models for the conditional (Eq. 10) and unconditional forecasts (Eq. 11). Results are based on forecasts made in 2002 for each year in the future. Note that conditional forecasts use future information on neighborhood growth rates. Shaded areas indicate twice the standard errors.

better every single year, with an average predictability gain of around 20%. In the supplementary information we investigate further aspects of the predictability gain distribution and present results for alternative forecasting benchmarks.

The results of the prediction exercises add further support to the analysis of the previous sections. Innovation rates are network-dependent, and having knowledge on the technology network helps improving forecasts of patenting dynamics.

5. Discussion

We have provided evidence that technological domains co-evolve if they are linked through the patent citation network. This empirical observation can be explained with a simple model of network-dependent knowledge creation. Motivated by the recursive nature of the innovation process, technologies take advantage of distinct knowledge sources in the technological ecosystem. We also validated the model by making out-of-sample prediction of growth rates, conditional and unconditional of neighboring growth rates. We have shown that the network model improves forecasts of patenting growth rates substantially when compared to standard time series models. Thus, the network of technological interdependencies is informative for inferring future patenting dynamics.

It is also important to point out the limitations and caveats of our analysis. By keeping the number of nodes fixed in the technological network, we ignored the emergence of breakthrough technologies. In our approach, technological domains are taken as given and well-defined by

the latest technology classification system. But technological classifications themselves are subject to evolutionary forces emerging from the changing technological base (Lafond and Kim, 2019). New technology classes are created, old ones merge or are abandoned and innovations can be reclassified – mechanisms which were not made explicit in this work.

In our empirical framework, we allowed the technology network to vary in time, but we did not model this explicitly. An interesting avenue for future research is to gain a better understanding of the mechanisms driving these rewiring processes. Furthermore, it would be important to understand how patenting activity and technology network metrics such as centrality translate into cost reductions of technologies (Farmer and Lafond, 2016; Triulzi et al., 2018). Finally, another interesting direction of research is to further investigate the interaction of network-dependent innovation dynamics with the real economy by coupling innovation network with input-output networks. This could illuminate new aspects of how structural change happens in the economy and improve economic forecasts.

We conclude by discussing the policy implications of our analysis. Since innovation dynamics in technological domains reveal strong network dependence, the allocation of research resources has to consider the ecosystem in which the focal domain is embedded in. Innovation in a domain is not an isolated process, but results from complex mechanisms involving research efforts, institutional settings, and technological interdependencies. Facing the enormous challenge of transitioning the current economic system to a low-carbon economy, we will need substantial improvements in key low-carbon technologies such as photovoltaics and wind energy. Better understanding their technological interdependencies emerging from the technological ecosystem and making it fruitful for research policy will be a crucial step in this direction.

Acknowledgments

We are grateful to Christian Diem, Torsten Heinrich, Giorgio Triulzi and the INET Complexity Economics research group for helpful discussions. This work was supported by Baillie Gifford, the Oxford Martin Programme on the Post-Carbon Transition, European Union’s Horizon 2020 research and innovation programme under grant agreement No. 730427 (COP21 RIPPLES) and Partners for a New Economy. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract no. 2019-1902010003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

- Acemoglu D, Akgigit U, Kerr WR. Innovation network. *Proceedings of the National Academy of Sciences* 2016;113(41):11483–8. doi:10.1073/pnas.1613559113.
- Aghion P, Howitt P. A model of growth through creative destruction. *Technical Report*; National Bureau of Economic Research; 1990. doi:10.3386/w3223.
- Alstott J, Triulzi G, Yan B, Luo J. Mapping technology space by normalizing patent networks. *Scientometrics* 2017;110(1):443–79. doi:10.1007/s11192-016-2107-y.
- Anselin L. *Spatial econometrics: methods and models*. volume 4. Springer Science & Business Media, 2013.
- Armstrong JS. *Long-Range Forecasting: From Crystal Ball to Computer*. volume 2. Wiley, 1985.
- Arthur WB. *The nature of technology: What it is and how it evolves*. Simon and Schuster, 2009.
- Bagley RJ, Farmer JD, Fontana W. Evolution of a metabolism. *Artificial life II* 1992;10:141–58.
- Cowan R, Jonard N. The dynamics of collective invention. *Journal of Economic Behavior & Organization* 2003;52(4):513–32. doi:10.1016/S0167-2681(03)00091-X.
- Elhorst JP. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. volume 479. Springer, 2014.
- Ertur C, Koch W. Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics* 2007;22(6):1033–62. doi:10.1002/jae.963.
- Farmer JD, Kauffman SA, Packard NH. Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena* 1986;22(1-3):50–67.
- Farmer JD, Lafond F. How predictable is technological progress? *Research Policy* 2016;45(3):647–65. doi:10.1016/j.respol.2015.11.001.
- Fink T, Reeves M. How much can we influence the rate of innovation? *Science Advances* 2019;5(1):eaat6107. doi:10.1126/sciadv.aat6107.
- Fleming L. Recombinant uncertainty in technological search. *Management Science* 2001;47(1):117–32. doi:10.1287/mnsc.47.1.117.10671.
- Grossman GM, Helpman E. Quality ladders in the theory of growth. *The Review of Economic Studies* 1991;58(1):43–61. doi:10.2307/2298044.
- Hall BH, Mairesse J, Mohnen P. Measuring the Returns to R&D. In: *Handbook of the Economics of Innovation*. Elsevier; volume 2; 2010. p. 1033–82. doi:10.1016/S0169-7218(10)02008-3.
- Ho CY, Wang W, Yu J. International knowledge spillover through trade: A time-varying spatial panel data approach. *Economics Letters* 2018;162:30–3. doi:10.1016/j.econlet.2017.10.015.
- Jain S, Krishna S. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proceedings of the National Academy of Sciences* 2001;98(2):543–7. doi:10.1073/pnas.98.2.543.
- Jones CI. R&D-based models of economic growth. *Journal of Political Economy* 1995;103(4):759–84. doi:10.1086/262002.
- Kauffman SA. *The origins of order: Self-organization and selection in evolution*. OUP USA, 1993.
- König MD, Battiston S, Napolitano M, Schweitzer F. Recombinant knowledge and the evolution of innovation networks. *Journal of Economic Behavior & Organization* 2011;79(3):145–64. doi:10.1016/j.jebo.2011.01.007.
- Lafond F, Kim D. Long-run dynamics of the US patent classification system. *Journal of Evolutionary Economics* 2019;1–34doi:10.1007/s00191-018-0603-3.
- Leontief WW. Quantitative input and output relations in the economic systems of the United States. *The Review of Economics and Statistics* 1936;:105–25doi:10.2307/1927837.
- Masuda N, Lambiotte R. *A Guide to Temporal Networks*. World Scientific, 2016.
- McNerney J, Farmer JD, Redner S, Trancik JE. Role of design complexity in technology improvement. *Proceedings of the National Academy of Sciences* 2011;108(22):9008–13. doi:10.1073/pnas.1017298108.
- Miller RE, Blair PD. *Input-output analysis: foundations and extensions*. Cambridge University Press, 2009.
- Napolitano L, Evangelou E, Pugliese E, Zeppini P, Room G. Technology networks: the autocatalytic origins of innovation. *Royal Society Open Science* 2018;5(6):172445. doi:10.1098/rsos.172445.
- Ord K. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 1975;70(349):120–6.
- Romer D. *Advanced Macroeconomics*. New York: McGraw-Hill, 2012.
- Romer PM. Endogenous technological change. *Journal of Political Economy* 1990;98(5/2):71–102. doi:10.1086/261725.
- Schumpeter JA. *Business Cycles*. New York: McGraw-Hill Book Company, 1939.
- Taalbi J. Evolution and structure of technological systems-an innovation output network. arXiv preprint arXiv:181106772 2018;.
- Tria F, Loreto V, Servedio VDP, Strogatz SH. The dynamics of correlated novelties. *Scientific Reports* 2014;4:5890. doi:10.1038/srep05890.
- Triulzi G, Alstott J, Magee CL. Estimating technology performance improvement rates by mining patent data 2018;doi:10.2139/ssrn.2987588.
- Usher AP. *A History of Mechanical Invention*. Cambridge, MA: Harvard University Press, 1954.
- Wang W, Yu J. Estimation of spatial panel data models with time varying spatial weights matrices. *Economics Letters* 2015;128:95–9. doi:10.1016/j.econlet.2015.01.021.
- Youn H, Strumsky D, Bettencourt LM, Lobo J. Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface* 2015;12(106):1–8. doi:10.1098/rsif.2015.0272.

Appendix A. Temporal technology network

The technology citation matrix C_t which quantifies knowledge spillovers between technologies at each point in time is not directly observable, but has to be derived from the patent citation matrix H_t . We first show for the static case how simple matrix algebra can be used to construct technology networks from patent data. Let us define the patent citation network H , where H_{pq} is one if patent p cites patent q and zero otherwise. To obtain the number of citations between technologies, the patent citation matrix, H , has to be projected onto the technology space. The patents-technology class relationship can be represented as a binary bipartite network where the link \tilde{B}_{pi} means that patent p belongs to technology class i . This setup is useful for deriving technology networks. Technological distances can be quantified in different ways, for example, with co-classification networks, co-citation networks or technology citation networks.

A simple one-mode projection of the bipartite network onto the set of CPC classes which does not take the citation matrix into account yields the co-classification network and is given by

$$C^{\text{co-classification}} = \tilde{B}^T \tilde{B}. \quad (\text{A.1})$$

The diagonal of $C^{\text{co-classification}}$ is then simply the count of patents in a given domain. As alternative measure of technological distance, the co-citation network can be derived by computing

$$C^{\text{co-citation}} = (\tilde{B}H)^T \tilde{B}H. \quad (\text{A.2})$$

The co-citation network quantifies how often two technology codes are cited together.

Since we are interested in the knowledge flows between technologies, the focus of this work is the technology citation network which can be obtained as

$$C^{\text{tech-citation}} = \tilde{B}^T H \tilde{B}. \quad (\text{A.3})$$

This projection yields a integer-valued citation matrix, but inflates the total number of citations in the technology-based network if multiple technology classes are assigned to a patent.

To keep total number of citations constant ($\sum_{ij} H_{ij} = \sum_{ij} C_{ij}$), we use the row normalized bipartite network in the projection. Note that the summing across rows $\sum_i \tilde{B}_{pi}$ yields the total number of CPC classes per patent p (the degree of p in the bipartite network) and the normalized entry $B_{pi} := \tilde{B}_{pi} / \sum_i \tilde{B}_{pi}$ is the “share” of technology i in patent p . Given the row-normalized bipartite network B , we obtain the technology citation matrix considered in the analysis by

$$C = B^T H B. \quad (\text{A.4})$$

Fig. A.4 illustrates the network projection in a schematic toy example.

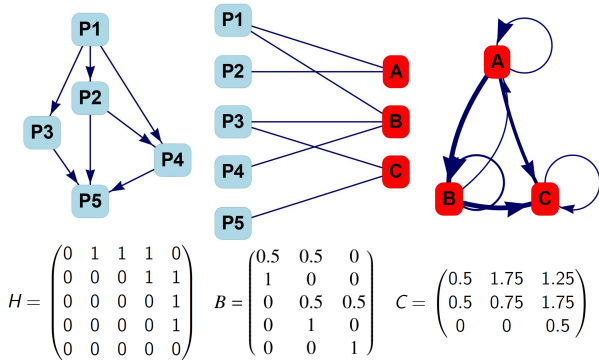


Figure A.4: A toy example of the one-mode projection in Eq. (A.4). The left network is a patent citation network H where a directed edge from i to j indicates that patent i cites j . The network is hierarchical as older patents cannot cite more recent patents. Below the network the corresponding adjacency matrix is shown. In the middle is the row-normalized bipartite patent-classification network B with some patents (blue) assigned to multiple technology classes (red). The projection yields the weighted and directed technology citation network C with self-loops shown on the right.

To obtain a temporal representation of knowledge flows, we index the patent citation network by time, H_t , where the time index refers to the citing patent. The time-indexed bipartite network B_t includes all the patents which cite and have been cited at time point t , as well as the corresponding technology classes. The time dependent technology citation matrix C_t is obtained by applying Eq. (A.4) to these matrices. The temporal network W_t row-normalizes the technology citation matrix as shown in the main text (Eq. (1)). The i^{th} row of W_t can be interpreted as the technological reliance of technology i at time t as it contains all the links from patents in i to all other technology classes at the particular time point. The structural

stability of the network is investigated further in the supplementary information.

Appendix B. Significance sampling

Technology citation networks are prone to be biased due to impinging factors such as the number of patents in a class or the age of patents. Size effects can drive the number of citations between technologies, because the expected number of citations between two technological classes increases with the number of patents in the classes. Another important source of bias is the age of a patent. A new patent might have few forward citations because other inventions simply did not have the time to cite the patent. To eliminate the size and age biases, Alstott et al. (2017) suggest an algorithm to sample random network realizations conditional on domain sizes and patent ages. The algorithm yields networks which preserve these impinging factors, but are random otherwise. The random realizations can then be used to derive a z-score for each link in the network. Links between technological domains which one would expect by chance, will have small z-scores and can be removed.

The randomized controls are generated as follows. First, identify all citations with a specific time lag which have been made in a particular year. An example would be to take all the citations made by patents in the year 2000 which cite patents three years earlier. Then, second, reshuffle all the cited patents. In the example, patents issued in 2000 would point randomly to the set of patents issued three years earlier. Larger classes will receive more citations on average. Note that the total number of citations for a given time lag are preserved by this procedure. By repeating the algorithm many times, an expected number of citations between classes can be derived, as well as the standard deviation. The link expectations and standard deviations are then used to compute the z-scores.

Sampling a large number of random networks for every year between 1947 and 2017 is computationally costly. Therefore, we analytically derive z-scores for every potential inter-domain link as suggested by Alstott et al. (2017) who have shown that the analytical derivation approximates the numerical randomization process very well. The randomization procedure can be stated as drawing from a hypergeometric distribution. Recall that a hypergeometric distribution is the probability of observing a certain number of successes in n draws, without replacement, from a sample of size s containing k objects with the success feature.

In our context, the hypergeometric distribution describes the probability of observing a certain number of citations from class i to class j at t with lag l . Let us denote the number of citations between two classes at t and given lag l with $C_{ij,t(t-l)}$. The sample size s is then simply the total number of inter-domain citations, $s_{t(t-l)} = \sum_{i,j} C_{ij,t(t-l)}(1 - \delta_{ij})$. The number of draws n corresponds to the number of citations made by a given class

i . Thus, $n_{i,t-l} = \sum_{j \neq i} C_{ij,t(t-l)}$. The number of successes k is the number of citations received by j at $t-l$, i.e. $k_{j,t-l} = \sum_{i \neq j} C_{ij,t(t-l)}$.

The expected number of patents from i to j at t for all possible lags can then be found by

$$\mathbb{E}[C_{ij,t}] = \sum_{l=0}^{t-1836} \frac{k_{j,t-l}}{s_{t-l}} n_{i,t-l}, \quad (\text{B.1})$$

and, by assuming independent random variables, the standard deviation is given by

$$\sigma_{ij,t} = \left(\sum_{l=0}^{t-1836} n_{i,t-l} \frac{k_{j,t-l}}{s_{t-l}} \left(1 - \frac{k_{j,t-l}}{s_{t-l}} \right) \frac{s_{t-l} - n_{i,t-l}}{s_{t-l} - 1} \right)^{1/2}. \quad (\text{B.2})$$

For each entry in the citation matrix we calculate z-scores based on

$$z_{ij,t} = \frac{C_{ij,t} - \mathbb{E}[C_{ij,t}]}{\sigma_{ij,t}}. \quad (\text{B.3})$$

We eliminate all citations in the technology citation matrix C_t with z-score smaller 2, roughly corresponding to the 97.5% quantile of the standard normal distribution. As we show in more detail in the supplementary information, the significance sampling yields substantially sparser networks.

Appendix C. Estimation of network model

The econometric network model, Eq. (6), is estimated using maximum likelihood since the OLS estimator is biased and inconsistent due to the reasons outlined in Anselin (2013). The model is related to spatial autoregressive (SAR) models for panel data. But in contrast to most other spatial econometric applications, our model uses a time-varying network term and includes diagonal elements which complicate the estimation. We restate the econometric model in matrix form which will prove useful in the derivation of the estimators.

The model and its corresponding data generating process can be written as

$$g_t = V_t a + \beta V_t \tilde{g}_t + \epsilon_t, \quad t \in \{1, 2, \dots, T\}, \quad (\text{C.1})$$

$$g_t = [\mathbb{I} - \beta W_t]^{-1} a + [\mathbb{I} - \beta W_t]^{-1} \epsilon_t, \quad (\text{C.2})$$

$$\epsilon_t \sim N(0, \sigma^2 \mathbb{I}),$$

where the matrix $V_t = V_t(\beta) \in \mathbb{R}^{N \times N}$ is diagonal with elements $V_{ii,t} := (1 - \beta W_{ii,t})^{-1}$. g_t , a , \tilde{g}_t and ϵ_t are column vectors of length N . The i^{th} element of \tilde{g}_t is defined as $\tilde{g}_{i,t} := \sum_{j=1}^N W_{ij,t} g_{j,t} (1 - \delta_{ij})$.

The likelihood function is given by

$$\begin{aligned} \text{LogL}((\sigma^2, a, \beta) | \{g_t\}) = & \quad (\text{C.3}) \\ -\frac{NT}{2} \ln(2\pi\sigma^2) + \sum_{t=1}^T \ln |\mathbb{I} - \beta V_t W_t| - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^\top \epsilon_t. \end{aligned}$$

To find the network specific parameter β , we first optimize the log-likelihood function with respect to σ^2 and a_i . Following the usual procedure in spatial autoregressive models, we then numerically optimize the concentrated log-likelihood function with respect to β (Ord, 1975; Elhorst, 2014; Wang and Yu, 2015). Maximizing the log-likelihood function with respect to σ^2 and a_i yields the following expressions:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_t \epsilon_t^\top \epsilon_t}{NT}, \\ \hat{a} &= \left(\sum_t V_t^2 \right)^{-1} \sum_t V_t g_t - \beta \left(\sum_t V_t^2 \right)^{-1} \sum_t V_t^2 \tilde{g}_t. \end{aligned}$$

By plugging the found expressions into Eq. (C.3), we find parameter β through the following optimization problem

$$\underset{\beta}{\text{argmax}} \quad \text{constant} + \sum_{t=1}^T \ln |\mathbb{I} - \beta V_t W_t| - \frac{NT}{2} \ln \sum_{t=1}^T \tilde{\epsilon}_t^\top \tilde{\epsilon}_t, \quad (\text{C.4})$$

with $\tilde{\epsilon}_t := g_t - V_t \left(\sum_t V_t^2 \right)^{-1} \sum_t V_t g_t - \beta V_t \left(\sum_t V_t^2 \right)^{-1} \sum_t V_t^2 \tilde{g}_t$. The (asymptotic) variance-covariance matrix is obtained as the Fisher information matrix, i.e. by inverting the numerically computed Hessian of the negative log-likelihood function.

Appendix D. Details on predictions

For both the ARIMA and the unconditional network model forecasts we fine-tune the models in-sample to achieve good out-of-sample forecasts of patenting levels. As frequently done in statistical learning, we split the data set into three parts: a training set (1948-1987), a validation set (1988-2002) and a test set (2003-2017). We first calibrate the models to the training set and use different specifications, as outlined below, to predict patenting in every domain in the validation set. We then choose the models which yield the lowest median absolute percentage error in the validation set forecasts. We use median absolute percentage errors over mean absolute percentage errors, since the well-known downward bias of the mean absolute percentage error (Armstrong, 1985) tends to favor predictors which are systematically underestimating. After choosing the best model specifications, we re-estimate the models to the whole data set up to 2002. Using these parametrizations of the ‘‘optimal’’ models, we then predict patenting activities for every single year and technology from 2003 to 2017.

In case of the ARIMA predictions, a choice has to be made regarding how many MA and AR terms to include in the predictions. Following the proposed principles, we first estimate all possible ARIMA models with lags $\{(p, q) | p, q \in \{0, \dots, 5\}\}$ for a given technology in the training set. Second, we use all estimated configurations to predict

patenting levels in the validation set. The model specification which reduces the median absolute percentage error in the validation set is then chosen to predict patenting activities in the test set. This procedure is repeated for every single time series.

For the unconditional network model forecasts we follow a similar approach to obtain the optimal order k' . To see why we optimize the network model with respect to k' , note that the data generating process in Eq. (C.2) for given parameters implies $\mathbb{E}[g_t] = (\mathbb{I} - \beta W_t)^{-1} a$ and $\mathbb{V}[g_t] = \sigma^2 [(\mathbb{I} - \beta W_t^\top)(\mathbb{I} - \beta W_t)]^{-1}$, entailing an exploding variance for β approaching one. Since our estimations suggest that β is actually close to one, forecast errors are expected to be large. To alleviate this issue, we take advantage of the power series expansion, $(\mathbb{I} - \beta W_t)^{-1} = \sum_{k=0}^{\infty} \beta^k W_t^k$, by defining the predictor as $\hat{g}_t^{uncond.} = (\sum_{k=0}^{k'} \hat{\beta}^k W_t^k) \hat{a}$. Increasing k' reduces the bias, versus reducing k' lowers the variance. To find the best choice of k' , we first estimate the model based on the training set and use the defined predictor to forecast patenting activities in the validation set. We then choose k' such that the median absolute percentage error, across time and technologies, in the validation set is minimized. The selection procedure yields $k' = 3$. As discussed in the main text, we found different magnitudes of network dependence for different growth rate lags. We therefore estimated the network model for every growth rate lag (1-year, 2-years, ..., 15-years) separately.

This approach assumes that the optimal choice in the validation set also corresponds to the best choice in the test set which does not always have to be the case. The dependence between variables can change over time, an aspect which is not taken into account here. Alternative approaches could be considered in future research. For the conditional forecasts, no further model fine-tuning has to be done.

Supplementary Information

Technological interdependencies predict innovation dynamics

Anton Pichler, François Lafond, J. Doyne Farmer

Contents

1	Data sources and preprocessing	2
2	Robustness to alternative classification systems	2
3	Temporal network stability	4
4	Nearest neighbor growth correlations	4
5	Econometric estimation	6
5.1	Econometric network model and SAR	6
5.2	Time-varying network model	6
5.3	Static network model	6
5.4	Time-varying spatial autoregressive model	7
6	Predicting innovation dynamics	9
6.1	ARIMA forecasts	9
6.2	Network model forecasts	10
6.3	Performance evaluation	10

1 Data sources and preprocessing

Our database contains all US utility patents granted between 1836 and 2018, except for rare exceptions. These patents are classified in two different classification systems: Cooperative Classification Scheme (CPC) and International Patent Classification (IPC) (up to 2015, they were also classified in the US Patent Classification System (USPC)).

CPC classifications are retrieved from the Master Classification File available from the USPTO/Reed Tech (version 2019-04-01). IPC classifications and patent dates are retrieved from Google Patents Public Data, provided by IFI CLAIMS Patent Services, accessed June 2018.

For the citation data, we merge the citation database of Kogan et al. (2017) with the citation data collected by Lafond and Kim (2019) and with the database provided by the USPTO (patentsview.org/download, accessed: 14/05/2019).

2 Robustness to alternative classification systems

We check the robustness of the results by applying our method to two different technology classification schemes, at two different levels in each case. In the main text, we discuss results based on the Cooperative Classification Scheme (CPC) 4-digits codes. Here, we also show results based on the CPC 3-digits, International Patent Classification (IPC) 3-digits and IPC 4-digits codes scheme. In the following, we describe the data in detail and how we have processed it for the presented analysis.

Fig. 1a) shows total patenting rates in the US from 1836 to 2018 for distinct classification codes of patents. While we find that CPC codes for almost every patent back to 1836 (the solid and the dotted line are almost indistinguishable in the plot), reclassification of IPC codes seems to be less rigorous for patents before 1920. For 2018 the data is not yet entirely updated, indicated by the sharp drop in patenting rates at the end of the time series. We therefore exclude the year 2018 in further analysis. Fig. 1b) shows the citations counts per year. The figure illustrates that reliable data on citations made is available only from 1947 on, but we have information on citations received by earlier patents.

We map patents onto their CPC and IPC codes and eliminate duplicates if the same patent was classified multiple times with the same 3- or 4-digits code. We only include technological classes which were at least ten times assigned to patents between 1836 and 2000. This removes no class at all in the CPC 3-digits case and only a single class in the CPC 4-digits case (A61P with only a single patent up to 2000). The cleaning is more effective in the IPC regimes where there are several classes which are very infrequently used. Introducing this threshold eliminates 143 IPC 3-digits and 389 IPC 4-digits classes. For the CPC classes, we also eliminate all Y-codes which are tags rather than separate technology classes. Table 1 summarizes some basic statistics of the used data, after excluding the year 2018.

Classification	# classes	# observations	# patents
CPC 3-digits	125	12,912,549	9,774,895
CPC 4-digits	649	14,422,099	9,774,895
IPC 3-digits	121	12,000,519	8,449,988
IPC 4-digits	633	13,697,093	8,449,720

Table 1: Descriptive statistics of used data after initial cleaning.

We obtain a temporal network for each of the four classification schemes. To avoid links between technological domains which one would expect by chance, we apply the significance sampling procedure (Alstott et al., 2017) described in the main text. In Fig. 2 we show the evolution of network density, i.e. the number of existing links divided by the number of all possible links (N^2), It becomes evident that the density is reduced substantially after applying the significance sampling procedure, but is also increasing in time. Unsurprisingly, the network density is higher when using the more aggregate 3-digits classification.

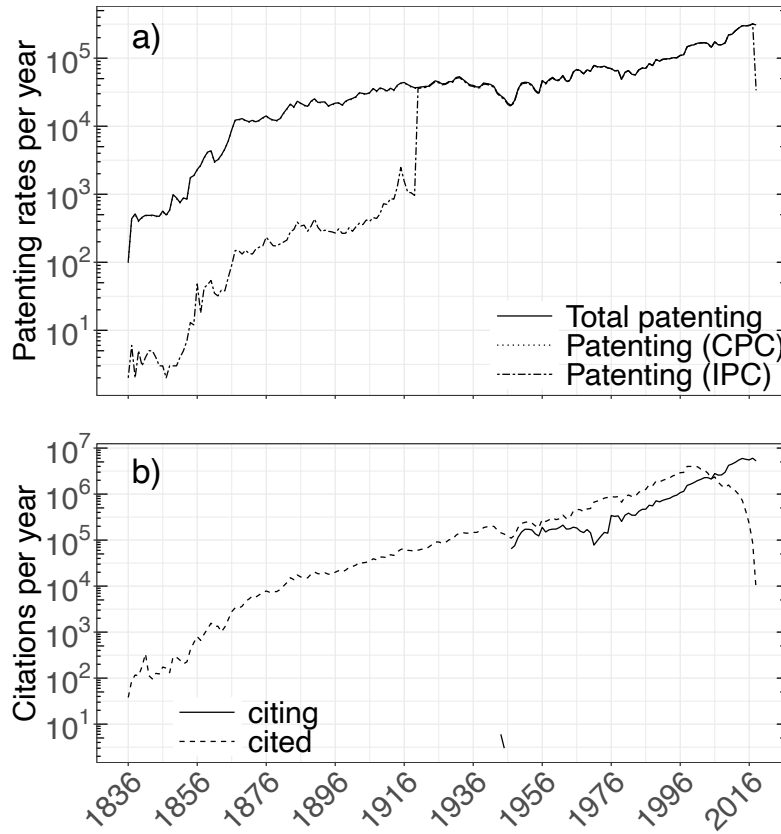


Figure 1: a) Yearly time series of US patents granted between 1836 and 2019. The solid line is the number of total granted patents, the dotted line is the number of patents with CPC attributes and the dotdashed line the number of patents having IPC attributes. b) Time series of citations per year in the data set. *Citing* refers to the number of citations made in a given year and *cited* refers to the number of citations received.

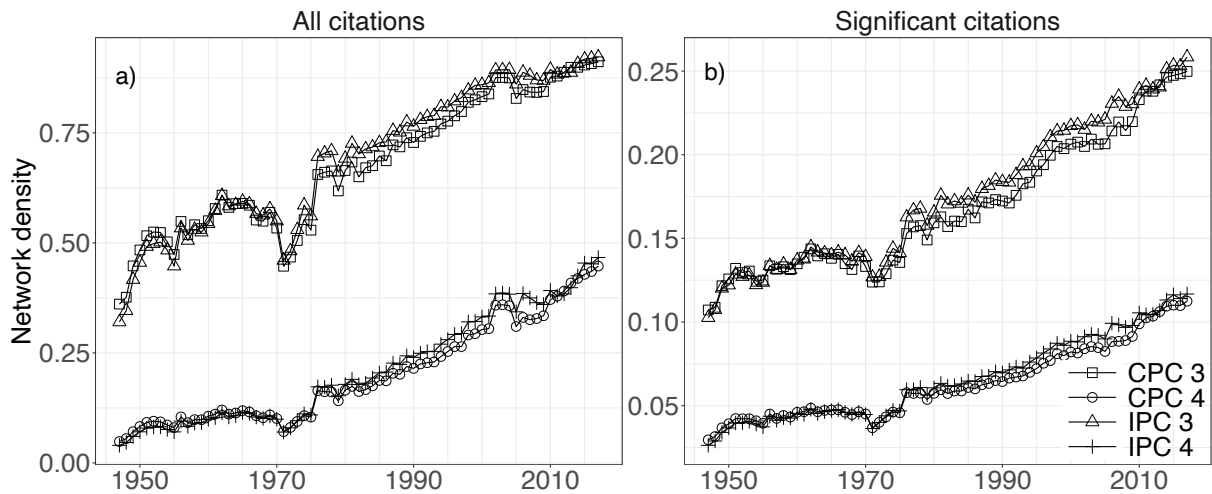


Figure 2: a) The network densities, number of existing links divided by the number of all possible links, over time for different classification schemes when all citations are included. b) The network densities after removing non-significant links.

3 Temporal network stability

As a measure of temporal stability, Fig. 3 shows the average cosine similarity between a row i of the technology matrix at t and $t - l$, or in more mathematical terms

$$\text{similarity}_{t,t-l}^{\text{av.}} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_j W_{ij,t} W_{ij,t-l}}{\sqrt{\sum_j W_{ij,t}^2} \sqrt{\sum_j W_{ij,t-l}^2}}.$$

We observe that the network tends to be fairly stable in time, i.e. the future technological dependence of a given technology is likely to be similar to its current one. After the temporal stability of the network was increasing for most parts of the second half of the last century, it became weaker in the last 20 years. As one would expect, we find that the average similarity decreases for higher lags, although there is still a pronounced positive relationship between technological dependences over a 10-year horizon. A qualitative difference between 3-digits and 4-digits classification schemes is apparent. The networks based on the more aggregated 3-digits classifications exhibit higher correlations in time for all shown lag choices. Moreover, the differences in temporal correlations between different time lags become less pronounced. This suggests that it is important to use more aggregated technology classification if the research is interested on the temporal evolution and structural change of the technological ecosystem.

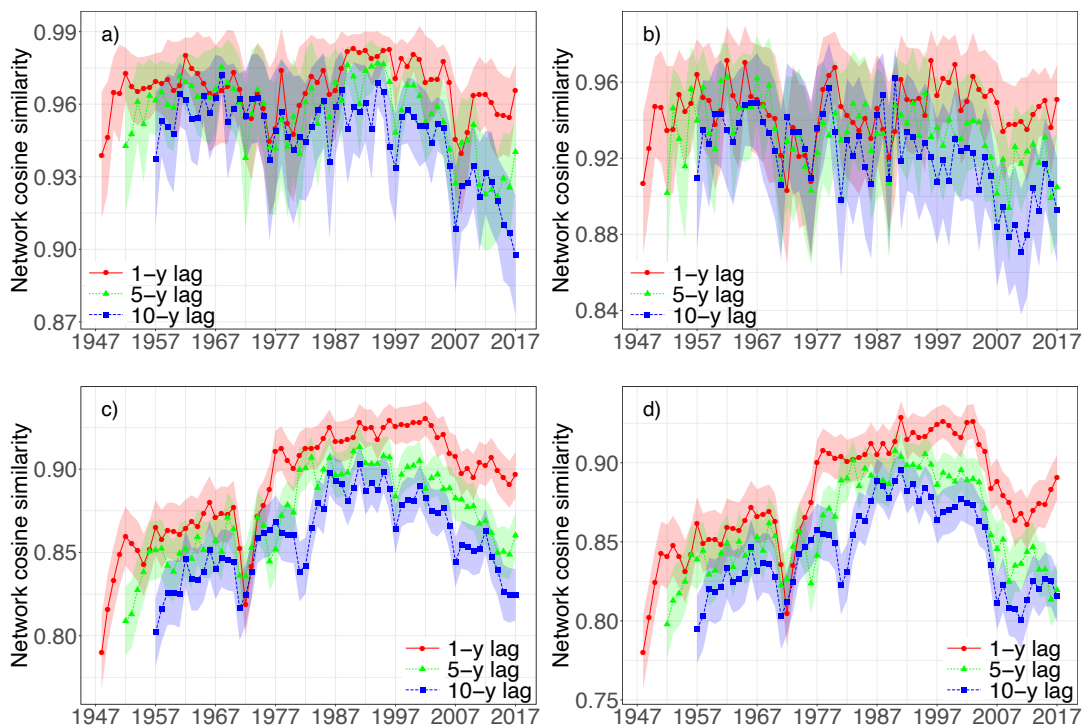


Figure 3: Average temporal cosine similarity of technology networks over time. The shaded area shows the range of the average plus/minus twice the standard errors. The results are shown for networks based on the a) 3-digits CPC codes, b) 3-digits IPC codes and c) 4-digits CPC codes and d) 4-digits IPC codes.

4 Nearest neighbor growth correlations

In the main text we have shown that the technology network is highly assortative with respect to growth rates, suggesting that patenting growth rates are similar if the technologies are connected. Fig. 4 plots the average nearest neighbor growth rates correlation over time for the differently aggregated networks. The positive assortativity pattern is more noisy in the more aggregated networks, panel a) and b), but still pronounced for most of the cases. The average nearest neighbor growth rates correlation is relatively insensitive with respect to different growth rate lengths choices for the more aggregated networks, again pointing at the importance of using more granular networks for discerning temporal effects in the technological evolution.

Table 2 summarizes the nearest neighbor growth correlations for different classifications and and lags. We see that the correlation magnitude is similar for different classification schemes, but systematically less noisy when using the more granular 4-digits classification. Data aggregation often reduces noise by averaging out extrema. Thus, it is not obvious a priori that the more fine-grained network representations exhibit less variation in their average nearest neighbor growth rates correlations.

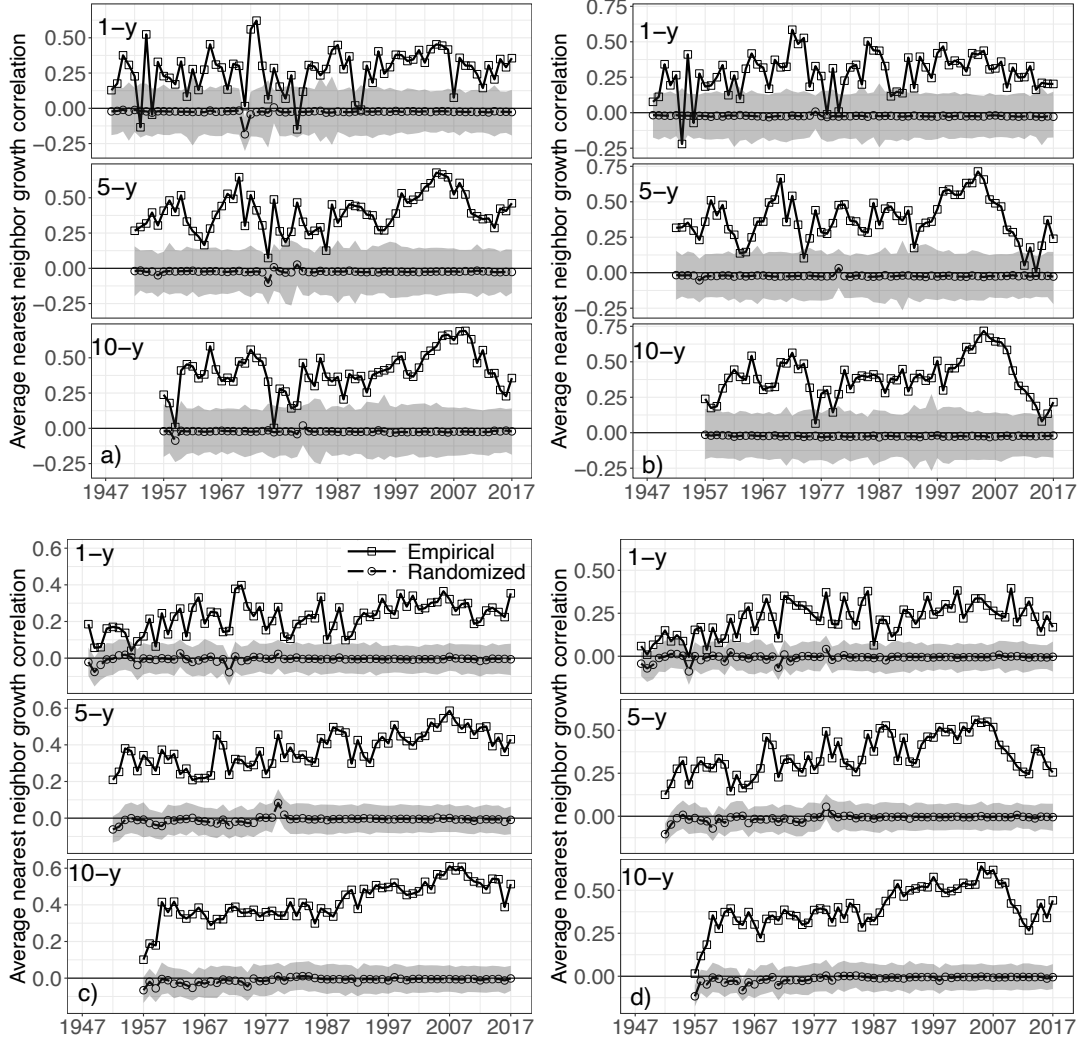


Figure 4: Average nearest neighbor growth rates correlation across time for networks based on the a) 3-digits CPC codes, b) 3-digits IPC codes, c) 4-digits CPC codes and d) 4-digits IPC codes. The shaded area is the 5-95% inter-quantile range of results based on 1,000 randomized networks (described in the main text).

	Lag	CPC 3	CPC 4	IPC 3	IPC 4
Time average	1-year	0.27	0.22	0.28	0.22
	5-year	0.39	0.38	0.38	0.35
	10-year	0.40	0.42	0.39	0.40
Time std. deviation	1-year	0.15	0.08	0.14	0.10
	5-year	0.13	0.10	0.15	0.11
	10-year	0.15	0.10	0.15	0.12

Table 2: Average nearest neighbor growth rate correlations. Time average and time standard deviation of the average nearest neighbor growth rates correlation for distinct technological classification schemes.

5 Econometric estimation

5.1 Econometric network model and SAR

In the main text we have presented results based on the econometric model

$$g_{i,t} = \frac{a_i}{1 - \beta W_{ii,t}} + \frac{\beta}{1 - \beta W_{ii,t}} \sum_{j=1}^N W_{ij,t} g_{j,t} (1 - \delta_{ij}) + \epsilon_{i,t}, \quad (1)$$

where $\epsilon_{i,t} \sim N(0, \sigma^2)$. Note that the econometric model is a generalization of the spatial autoregressive (SAR) model for panel data. In case of a zero diagonal matrix, $W_{ii} = 0$, the model reduces to the fixed-effects SAR with time-varying spatial component

$$g_{i,t} = a_i + \beta \sum_{j=1}^N W_{ij,t} g_{j,t} + \epsilon_{i,t}, \quad (2)$$

discussed by Wang and Yu (2015). Moreover, if the network is fixed in time, $W_t = W$, this model reduces to the standard fixed-effects SAR which can be estimated as outlined in Elhorst (2014).

5.2 Time-varying network model

We estimate different variations of these econometric models to see how robust the results in the main text are. First, we calibrate the model as discussed in the main text to all four classification schemes and report the results in Table 3. The results are qualitatively similar for different aggregations.

The econometric estimation also allows use to make the magnitude of network effects in patenting growth rates explicit. Let us define the average direct and average total impacts of research effort growth on patenting growth by

$$\bar{I}_{direct,t} := \alpha \frac{\sum_{i=1}^N L_{ii,t}}{N}, \quad (3)$$

$$\bar{I}_{total,t} := \alpha \frac{\sum_{i=1}^N \sum_{j=1}^N L_{ij,t}}{N}, \quad (4)$$

respectively, yielding the relative network impact of

$$I_{network,t} := 1 - \frac{\bar{I}_{direct,t}}{\bar{I}_{total,t}} = 1 - \frac{\sum_{i=1}^N L_{ii,t}}{\sum_{i=1}^N \sum_{j=1}^N L_{ij,t}}. \quad (5)$$

Fig. 5 shows the relative network impacts $I_{network,t}$ over time for all four classifications. The network effects are growing in time which is not surprising given the increase in network density we have observed. The figure shows that we would estimate higher network impacts in a more aggregated network, i.e. when based on 3-digits classifications. We also find that network effects are slightly larger for IPC codes than for CPC codes, if compared on the same digit-level. But the general trend of rising relative network impacts over time holds for all four classification schemes.

5.3 Static network model

We next investigate how the estimated parameters change if we use fixed network specifications. Patent networks are naturally dynamic, but the vast majority of the spatial econometrics literature does not allow for time-varying networks. We thus test whether our results also hold if we keep the technology network fixed.

We do this by estimating the model for every network snapshot separately, i.e. by setting $W_t = W$ in Eq. (1) where W represents a particular network in the sample. We then check how the parameter β changes if we use a different network snapshot instead. It should be pointed out that there might be an endogeneity issue if more recent networks are included in the model, since the network formation itself could depend on the growth of technological domains. Moreover, this estimation uses future information on the right-hand side of Eq. (1) to explain past left-hand side variables, for example, when the network of 2017 is used as W . Nevertheless, we estimate the model with a fixed, potentially future, network to see how results change from the time-varying to the static specification.

In Fig. 6 we plot the coefficient β as a function of the used network W_t . We show the results for CPC 3-digits and CPC 4-digits classes only, since results are similar for the IPC counterparts. The figure depicts qualitatively different results for the two aggregation levels. For the CPC 3-digits case, we find that β exhibits a slight downward trend as a function of the time index. In the CPC 4-digits case, on the other hand, the coefficient is smaller in earlier and larger for more recent networks. Additionally, from the mid-70s on the parameter seems to stabilize around the value of 0.84 which we have estimated in the original time-varying specification.

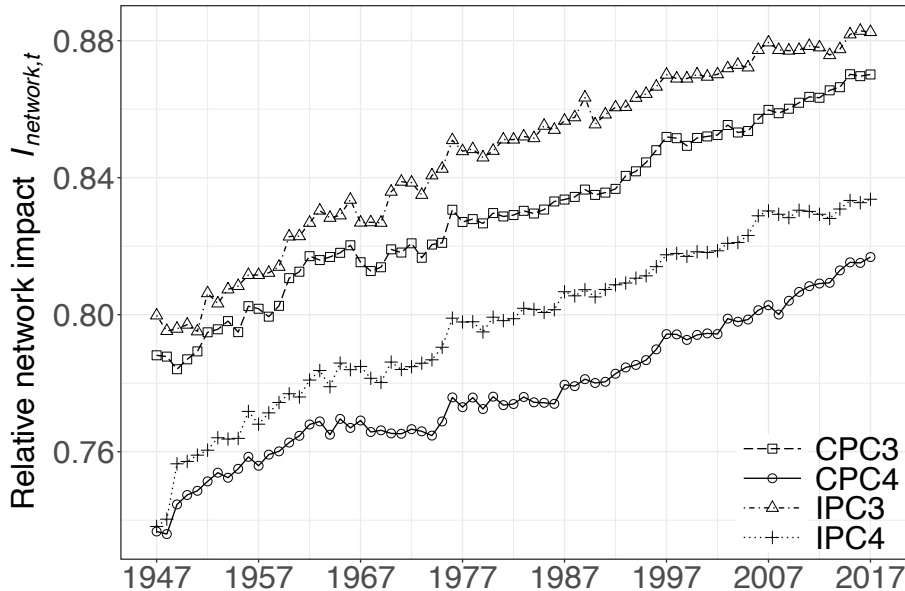


Figure 5: Relative network impacts over time for different classification schemes. Results are based on 1-year growth rates.

5.4 Time-varying spatial autoregressive model

We also estimated a standard SAR model (zero-diagonal) with time-varying network component (Eq. 2) for comparison. The SAR model is convenient for demonstration purposes, since extensive literature renders it a “standard” model for studying network effects. Due to its simpler form, it is also straightforward to include further regressors to check for robustness. We therefore estimate Eq. (2) also with additional regressors such as time effects and technological class sizes.

The estimation of the SAR model simplifies in comparison to the full network model of Eq. (1). To estimate Eq. (2) including exogenous regressors, we stack all elements $g_{i,t}$ into the NT -dimensional vector $\bar{g} := (g_{1,1}, g_{2,1}, \dots, g_{N,1}, g_{1,2}, \dots, g_{N,T})$. In the same manner we obtain the vector $\bar{\epsilon}$. By defining the T -dimensional vector of ones ι_T and the block-diagonal matrix $W := \text{diag}(W_1, \dots, W_T)$, we can rewrite

Dependent variable: patenting growth

	CPC3			CPC4			IPC3			IPC4		
	(1-y)	(5-y)	(10-y)	(1-y)	(5-y)	(10-y)	(1-y)	(5-y)	(10-y)	(1-y)	(5-y)	(10-y)
average \hat{a}_i	0.01	0.03	0.05	0.01	0.03	0.04	0.01	0.02	0.05	0.01	0.03	0.05
$\hat{\beta}$	0.92*** (0.006)	0.94*** (0.010)	0.94*** (0.015)	0.85*** (0.006)	0.93*** (0.006)	0.94*** (0.008)	0.92*** (0.006)	0.94*** (0.011)	0.92*** (0.019)	0.87*** (0.006)	0.93*** (0.007)	0.93*** (0.009)
$\hat{\sigma}^2$	0.03	0.08	0.13	0.11	0.20	0.30	0.03	0.09	0.14	0.11	0.23	0.31
Log-likelihood	2,525	-277	-364	-13,387	-5,474	-3,625	2,867	-384	-395	-13,431	-6,049	-3,677
Observations	8,610	1,720	858	43,651	8,734	4,348	8,349	1,669	834	42,688	8,528	4,427

Note:

*** $p < 10^{-16}$

Table 3: Results from estimating the econometric network model presented in the main text (Eq. 1) for different classification schemes. The results are shown for 1-, 5- and 10-year growth rates.

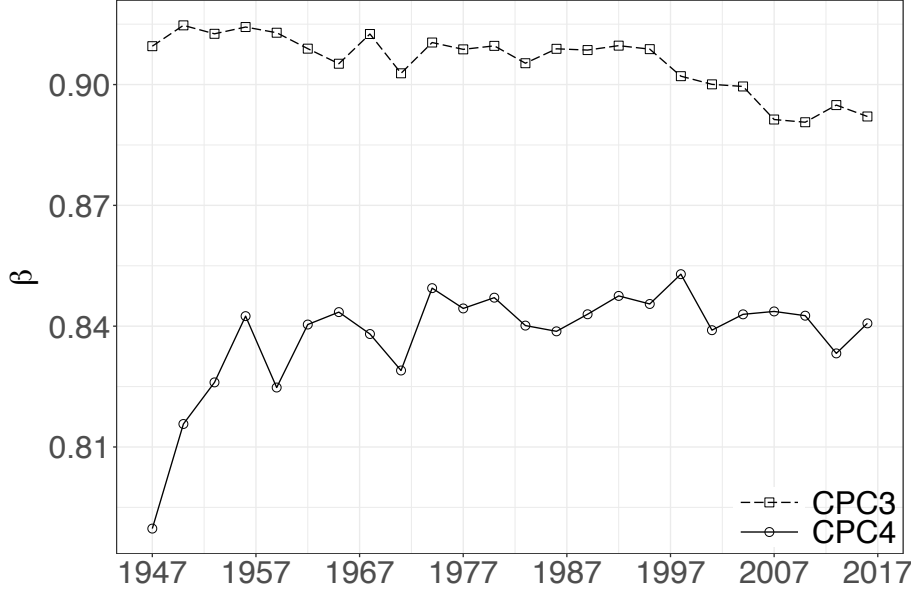


Figure 6: Parameter β obtained from estimating Eq. (1), but with a fixed instead of a time-varying network. The x-axis indicates the year of which the network snapshot was obtained from. For example, at 1947 the model was estimated with using $W_t = W_{1947}$ for all t . Results are based on 1-year growth rates.

Eq. (2) into matrix notation

$$\bar{g} = (\iota_T \otimes \mathbb{I}_N)a + \beta W \bar{g} + \bar{X} \delta + \bar{\epsilon}, \quad (6)$$

where \bar{X} is the $NT \times K$ matrix of exogenous regressors and δ the corresponding K -dimensional vector of coefficients. After subtracting the time average, Eq. (6) simplifies to

$$g = \beta W g + X \delta + \epsilon, \quad (7)$$

where g , X and ϵ are the time-demeaned versions of \bar{g} , \bar{X} and $\bar{\epsilon}$, respectively. The log-likelihood of the

<i>Dependent variable: patenting growth</i>										
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
	CPC3	CPC3	CPC3	CPC3	CPC3	CPC4	CPC4	CPC4	CPC4	CPC4
average \hat{a}_i	0.011	0.023	0.428	0.400	0.173	0.015	0.024	0.560	0.357	0.127
$\hat{\beta}$	0.830*** (0.009)	0.509*** (0.018)	0.513*** (0.017)	0.514*** (0.018)	0.511*** (0.018)	0.756*** (0.008)	0.483*** (0.105)	0.476*** (0.09)	0.483*** (0.105)	0.482*** (0.105)
Patent rates			-0.069*** (0.003)					-0.128*** (0.002)		
Patent stock				-0.038*** (0.004)					-0.041*** (0.003)	
Discount. pat. stock					-0.033*** (0.004)					-0.035*** (0.003)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Log-likelihood	2,527	2,721	2,912	2,763	2,757	-13,374	-12,923	-11,452	-12,854	-12,864
Observations	8,601	8,601	8,601	8,601	8,601	43,651	43,651	43,651	43,651	43,651

Note:

*** $p < 10^{-16}$

Table 4: Results from estimating the SAR with time-varying network component and additional exogenous regressors (Eq. 6) for different classification schemes. The first five columns (a-e) show the estimated parameters for the CPC 3-digits aggregation and the last five columns (f-j) for the CPC 4-digits aggregation. The results are based on 1-year growth rates.

time-demeaned model then reads

$$\text{LogL} = -\frac{NT}{2} \ln(2\pi\sigma^2) + \sum_{t=1}^T \ln |\mathbb{I} - \beta W_t| - \frac{1}{2\sigma^2} \epsilon^\top \epsilon, \quad (8)$$

with $\epsilon = (\mathbb{I}_{NT} - \beta W)g - X\delta$. The parameter β is obtained by solving

$$\begin{aligned} \underset{\beta}{\text{argmax}} \quad & \text{constant} + \sum_{k=1}^{NT} \ln(1 - \beta e_k) \\ & - \frac{NT}{2} \ln(g^\top g - 2\beta g^\top Wg - \beta^2 g^\top W^\top Wg - A^\top X(X^\top X)^{-1} X^\top Ag), \end{aligned}$$

where $A := (\mathbb{I}_{NT} - \beta W)$ and e_k represents eigenvalues of matrix W . Under standard regularity assumptions, we derive the variance-covariance matrix by inverting the Fisher information matrix which yields

$$\text{var}(\sigma^2, \beta, \delta) = \begin{bmatrix} \frac{NT}{\sigma^4} & -\mathbb{E} \left[\frac{\partial^2 \text{LogL}}{\partial \sigma^2 \partial \beta} \right] & 0_K^\top \\ \cdot & -\mathbb{E} \left[\frac{\partial^2 \text{LogL}}{\partial \beta^2} \right] & \frac{1}{\sigma^2} X^\top W A^{-1} X \delta \\ \cdot & \cdot & \frac{1}{\sigma^2} X^\top X \end{bmatrix}^{-1}, \quad (9)$$

where $-\mathbb{E} \left[\frac{\partial^2 \text{LogL}}{\partial \delta^2} \right] = \text{trace}(A^{-1} W A^{-1} W) + \frac{1}{2\sigma^2} \text{trace}([A^{-1} X \delta \delta^\top X^\top A^{\top-1} + \sigma^2 A^{-1} A^{\top-1}] W^\top W)$ and $-\mathbb{E} \left[\frac{\partial^2 \text{LogL}}{\partial \sigma^2 \partial \beta} \right] = \frac{1}{\sigma^2} \text{trace}(A^\top W [A^{-1} X \delta \delta^\top X^\top A^{\top-1} + \sigma^2 A^{-1} A^{\top-1}] + \delta^\top X^\top W A^{-1} X \delta)$. 0_K denotes a K -dimensional vector of zeros. Matrix elements indicated by the dots are not reported since they are trivially filled due to symmetry.

In Table 4 we show the results of estimating the SAR with time-varying network component for the CPC 3-digits, columns (a) to (e), and CPC 4-digits aggregation, columns (f) to (j). We find that the network coefficient β is estimated to be slightly larger, if the more aggregated networks are used. We also include time effects in the regression which reduces the network parameter in the CPC 4-digits case from 0.76 to roughly 0.48. When estimating the model for higher growth rate lags (not shown here), including time effects has less impact on the coefficient. The coefficient is fairly robust against adding further regressors such as including logged patenting rates, columns (c) and (h), logged cumulative patent stock, (d) and (i), or logged discounted cumulative patent stock, (e) and (j), where we have used a 15% yearly discount factor as suggested by Hall et al. (2005). The exogenous regressors are highly significant and have a negative sign, indicating that higher patenting rates and larger knowledge stocks at time point t come with smaller patenting rate growth in the following year. The average magnitude of the fixed effects increases as a consequence of including exogenous regressors.

6 Predicting innovation dynamics

In the main text we have shown that predictions of patenting rates in technological classes can be significantly improved if network effects are taken into account. In this section we present further details on predicting patenting rates with the introduced model. Results shown in the main text are based on the CPC 4-digits codes. Here, we show how the model predictions change when using alternative aggregation levels. Moreover, we discuss further aspects of the presented results and investigate alternative forecast benchmarks.

6.1 ARIMA forecasts

We first focus on the forecasts obtained from the ARIMA models which we state again for the ease of reference:

$$g_{i,t} = \nu_i + \sum_{s=0}^p \phi_{i,s} g_{i,t-s} + \sum_{s=0}^q \psi_{i,s} u_{i,t-s} + u_{i,t}, \quad (10)$$

where $\phi_{i,0} = \psi_{i,0} = 0$. As indicated in the main text, we have fitted every (p, q) combination up to order (5,5) of the ARIMA($p, 1, q$) model for each (logged) time series in the training set (1947-1987) and used the parameter fits to forecast the time series in the validation set (1988-2002). We then refitted the ARIMA($p, 1, q$) which yielded the smallest median absolute percentage error in the validation set to the data up to 2002 to forecast the test set from 2003 to 2017. Table 5 gives an overview of the best

(p,q) combinations for different aggregation methods. In the CPC 4-digit case discussed in the main text, almost 25% of model specifications use less than two AR and MA terms, with the the geometric random walk ($p = q = 0$) alone yielding the best forecasts already in 12% percent of all cases. But we also find time series where higher-order models are preferred. Similarly for alternative aggregations, the geometric random walk model is the most frequent choice.

6.2 Network model forecasts

To obtain the unconditional forecasts, a similar model selection procedure is applied. Here, we first calibrate the network model to the data in the training set and choose, as discussed in the main text, the parameter k' such that the median absolute percentage error is minimized in the validation set. The model is then refitted to the data, including training and validation set (1947-2002), to predict patenting in the test set from 2003 to 2017 with given k' . When doing this for the four different technology aggregations, we obtain the following results: $k' = 9$ for CPC 3-digits, $k' = 3$ for CPC 4-digits, $k' = 6$ for IPC 3-digits and $k' = 2$ for IPC 4-digits. Thus, the applied model selection procedure always prefers $k' > 0$, i.e. the procedure always suggests to include network effects in the predictions.

6.3 Performance evaluation

As in the main text, we compare forecasts based on the predictability gain measure

$$PG1_{i,t} = \frac{|P_{i,t} - \hat{P}_{i,t}^{ARIMA}| - |P_{i,t} - \hat{P}_{i,t}^{network}|}{|P_{i,t}|}, \quad (11)$$

i.e. taking the difference between ARIMA and network model absolute percentage errors (APE) for every predicted value. We also check the predictability gains based on the absolute errors (AE) instead of the absolute percentage errors by computing

$$PG2_{i,t} = |P_{i,t} - \hat{P}_{i,t}^{ARIMA}| - |P_{i,t} - \hat{P}_{i,t}^{network}|. \quad (12)$$

We then take averages, standard deviations, medians and interquartile ranges for every year and plot the results for all four aggregation schemes in Fig. 7 - 10. In every figure, panel a) shows the average predictability gains based on APE, Eq. (11). Panel b) gives the corresponding yearly median values and interquartile ranges of the APE-based predictability gains. Averages and standard errors of AE predictability gains, Eq. (12), are shown in panel c) and the corresponding medians and interquartile ranges are visualized in panel d). We discuss the results for each aggregation scheme separately.

CPC3		MA order q					
		0	1	2	3	4	5
AR order p	0	0.15	0.01	0.03	0.03	0.02	0.02
	1	0.04	0.00	0.01	0.00	0.02	0.02
	2	0.03	0.02	0.02	0.03	0.02	0.04
	3	0.03	0.02	0.03	0.00	0.02	0.02
	4	0.02	0.02	0.01	0.02	0.03	0.07
	5	0.02	0.02	0.01	0.03	0.05	0.07
IPC3		MA order q					
		0	1	2	3	4	5
AR order p	0	0.10	0.03	0.04	0.02	0.03	0.02
	1	0.01	0.03	0.02	0.00	0.03	0.03
	2	0.08	0.00	0.02	0.01	0.01	0.04
	3	0.03	0.00	0.02	0.04	0.02	0.03
	4	0.04	0.03	0.01	0.03	0.03	0.03
	5	0.01	0.05	0.02	0.02	0.08	0.05
CPC4		MA order q					
		0	1	2	3	4	5
AR order p	0	0.12	0.04	0.03	0.02	0.03	0.03
	1	0.06	0.02	0.01	0.02	0.02	0.04
	2	0.02	0.02	0.02	0.03	0.03	0.03
	3	0.03	0.01	0.02	0.01	0.03	0.03
	4	0.03	0.02	0.03	0.01	0.03	0.03
	5	0.03	0.03	0.01	0.03	0.03	0.02
IPC4		MA order q					
		0	1	2	3	4	5
AR order p	0	0.13	0.05	0.02	0.02	0.02	0.04
	1	0.05	0.02	0.01	0.02	0.02	0.04
	2	0.02	0.02	0.02	0.02	0.03	0.03
	3	0.02	0.00	0.01	0.03	0.02	0.03
	4	0.03	0.01	0.01	0.02	0.03	0.03
	5	0.04	0.04	0.03	0.03	0.04	0.03

Table 5: Distribution of (p, q) pairs which minimize the median absolute percentage error in the validation set for the ARIMA($p,1,q$) forecasts. p is the number of autoregressive terms and q the number of moving average terms in the ARIMA model. For example, in the CPC3 case we have 0.15 for the combination $(p, q) = (0, 0)$, meaning that for 15% of all cases the geometric random walk with drift yielded the best predictions in the validation set.

CPC 3-digit codes (Fig. 7):

Similarly as the results shown in the main text, predictions can be substantially improved in the conditional forecasts where the growth rates of a focal technology’s neighborhood is known. This holds true for averages and medians, regardless of the considered performance metric. In contrast to the result in the main text, we find that the unconditional forecasts are not able to beat the ARIMA model systematically. The main reason why in this scenario the unconditional predictions underperform is that the optimal tuning parameter k' is not stable between the validation and test set. While a large $k' = 9$ is chosen based on the validation set forecasts, this choice is suboptimal when predicting patenting rates in the test set. This is largely a result of temporal instability of coefficients. Estimating the model from earlier years yields higher fixed effects and a slightly lower β than when estimating it from more recent data, leading to an overprediction of patenting levels in the test set. A potential way to tackle this problem in future research is to allow for time-varying coefficients.

CPC 4-digit codes (Fig. 8):

In all four panels predictability gains of conditional forecasts are substantial. There is more variation in the unconditional forecasts. Panel a) is the same plot as presented in the main text which plots average $PG1$ (solid lines) and twice the standard errors (shaded areas) over time. In both prediction exercises, the network model comes with substantial positive predictability gains. In panel b) the median predictability gains, as well as the interquartile range are shown. In the unconditional forecasts, we see positive median predictability gains in the first eleven years, ranging from 1.3% to 30% and negative predictability gains in the last four years (−5% to −13%) when using $PG1$. The conditional forecasts always exhibit positive median predictability gains. Panels c) shows the average predictability gains based on absolute errors. On average, the network model outperforms the ARIMA forecasts in both prediction exercises, with the conditional forecasts yielding positive prediction gains in every year and the unconditional forecasts in 11 out of 15 years. Panel d) plots the median predictability gains based on the corresponding interquartile range for $PG2$ which yields qualitatively similar results as panel b).

IPC 3-digit codes (Fig. 9):

The average predictability gains over the ARIMA models in the conditional forecast setting reach almost 60% in 2009 when measured in percentages, panel a), or roughly 800 patents when looking at absolute errors, panel b). In the unconditional scenario, average predictability gains based on $PG1$ are less pronounced (although positive in every single year), but are substantial when considering the absolute error based $PG2$ measure. The median predictability gain based on the absolute percentage error in the conditional forecasts scenario lies between 1.6% and 55% as can be seen in panel c). In the unconditional scenario, the median gains are smaller, but always positive. Panel d) reveals similar patterns as the ones observed in panel c).

IPC 4-digit codes (Fig. 10):

For the IPC 4-digit codes we find average percentage predictability gains between 4% in 2003 and 74% in 2013 when considering the conditional network forecasts; see panel a). The predictability gains for the unconditional forecasts can get as high as 29% in the year 2016. Panel b) shows that the median predictability gains are always positive for the conditional forecasts (ranging from 5% to 68%) and positive for the unconditional forecasts in 13 out of 15 forecasts. Only in the last two years, the median predictability gains are slightly negative (less than one percent). When benchmarking the forecasts with absolute errors, Eq. (12), we find that predictability gains are always positive on average and tend to increase in time for both prediction scenarios; panel c). Qualitatively, panel d) resembles panel b) where median predictability gains are always positive, except for the last two years in the unconditional forecasts.

Figure 7: Prediction results based on **CPC 3-digit codes**. a) Average predictability gain based on $PG1$, Eq. (11). Shaded areas indicate twice the standard error. b) Median predictability gain based on $PG1$. Shaded areas indicate the interquartile range. c) Average predictability gain based on $PG2$, Eq. (12). Shaded areas indicate twice the standard error. d) Median predictability gain based on $PG2$. Shaded areas indicate the interquartile range.

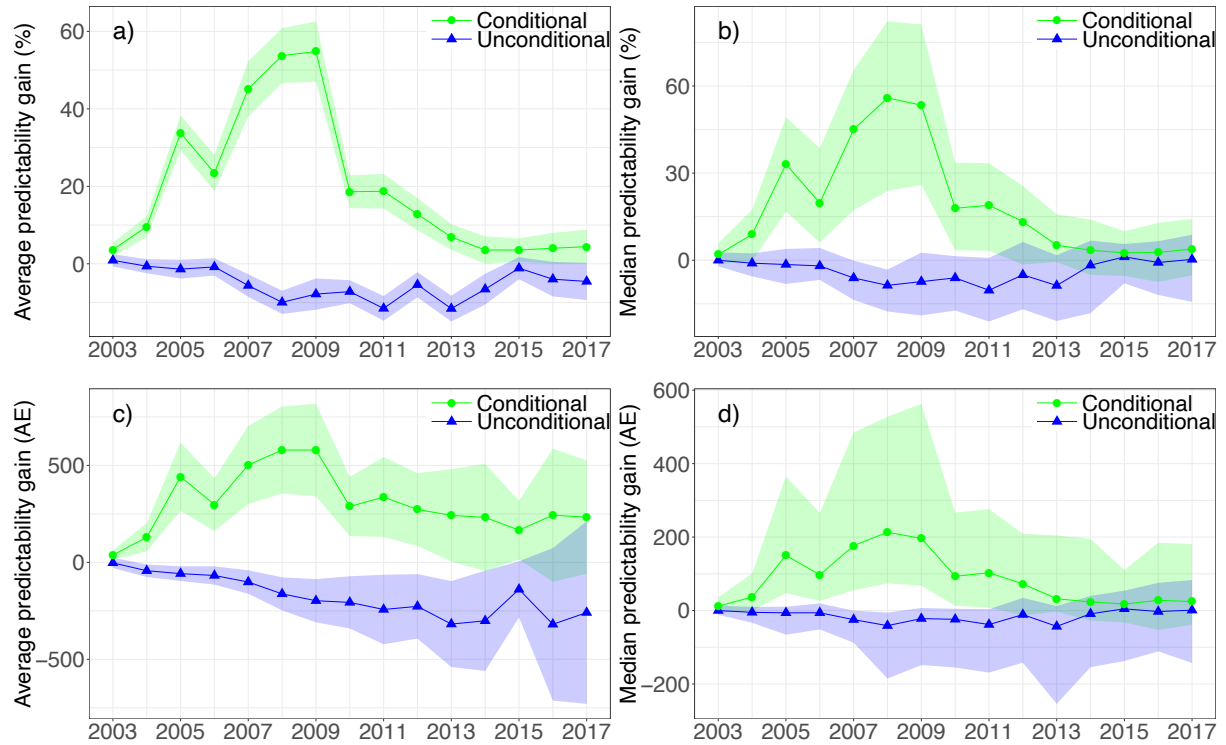


Figure 8: Prediction results based on **CPC 4-digit codes**. a) Average predictability gain based on $PG1$, Eq. (11). Shaded areas indicate twice the standard error. b) Median predictability gain based on $PG1$. Shaded areas indicate the interquartile range. c) Average predictability gain based on $PG2$, Eq. (12). Shaded areas indicate twice the standard error. d) Median predictability gain based on $PG2$. Shaded areas indicate the interquartile range.

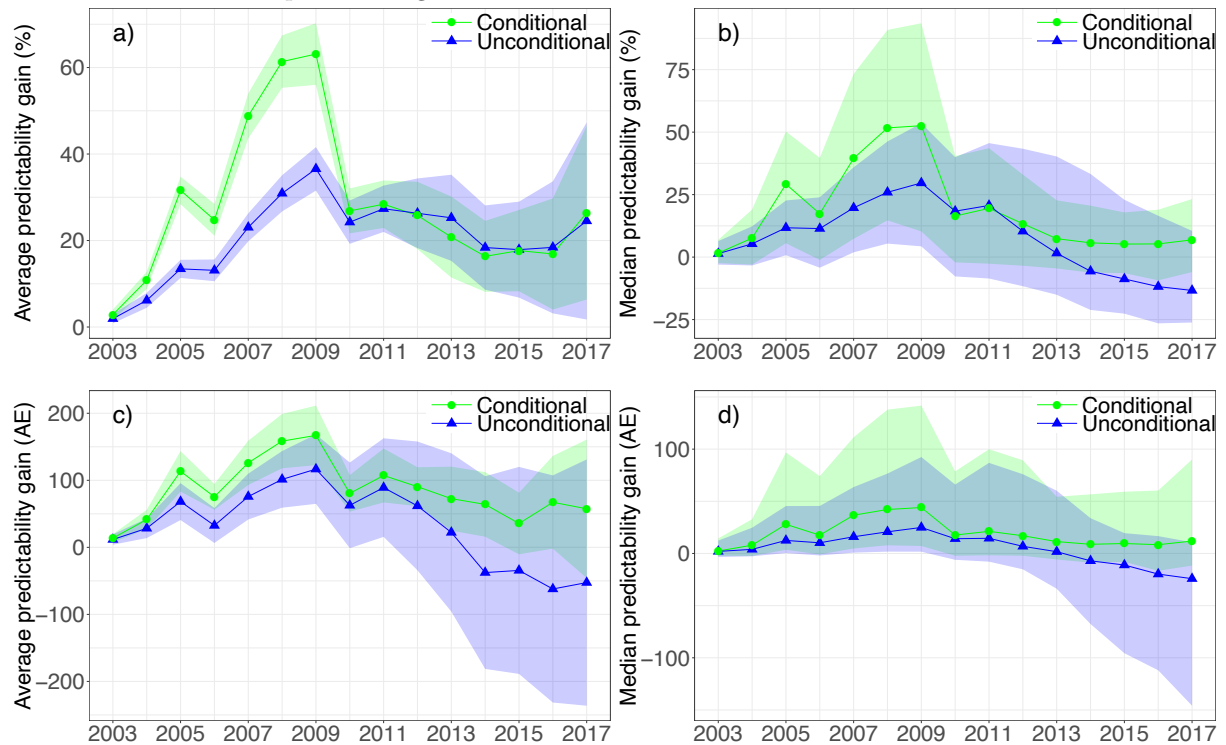


Figure 9: Prediction results based on **IPC 3-digit codes**. a) Average predictability gain based on $PG1$, Eq. (11). Shaded areas indicate twice the standard error. b) Median predictability gain based on $PG1$. Shaded areas indicate the interquartile range. c) Average predictability gain based on $PG2$, Eq. (12). Shaded areas indicate twice the standard error. d) Median predictability gain based on $PG2$. Shaded areas indicate the interquartile range.

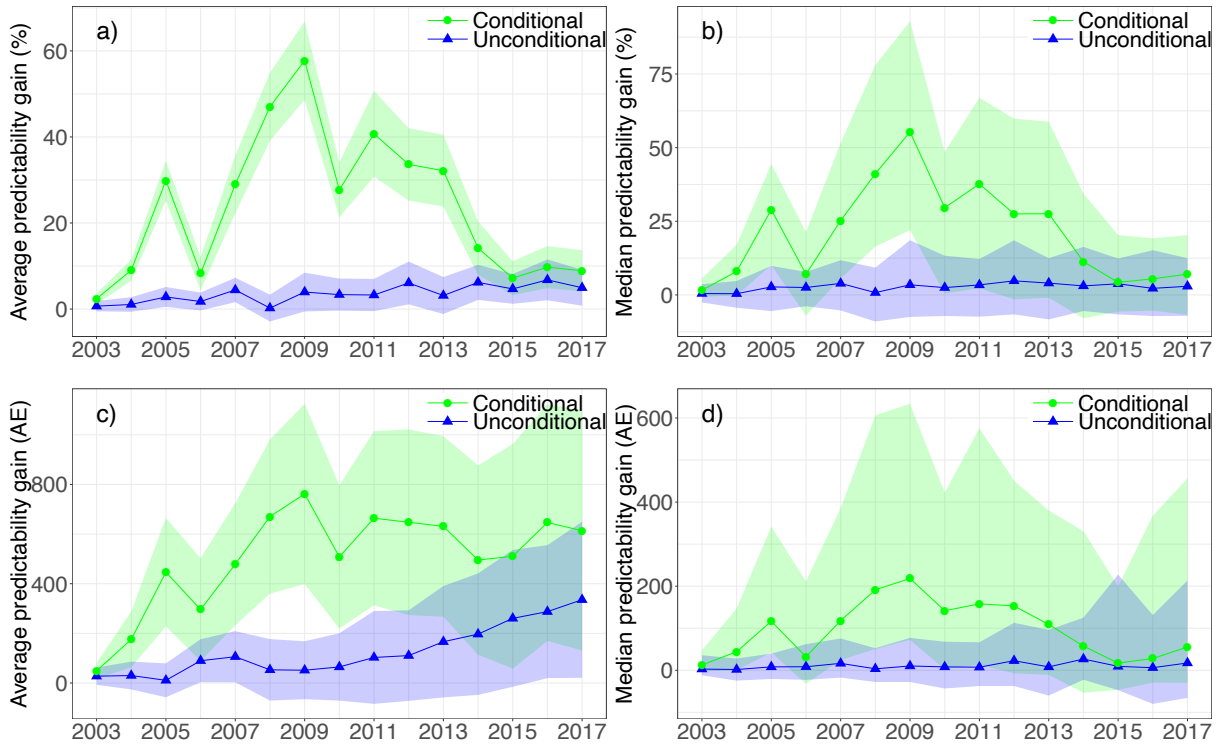
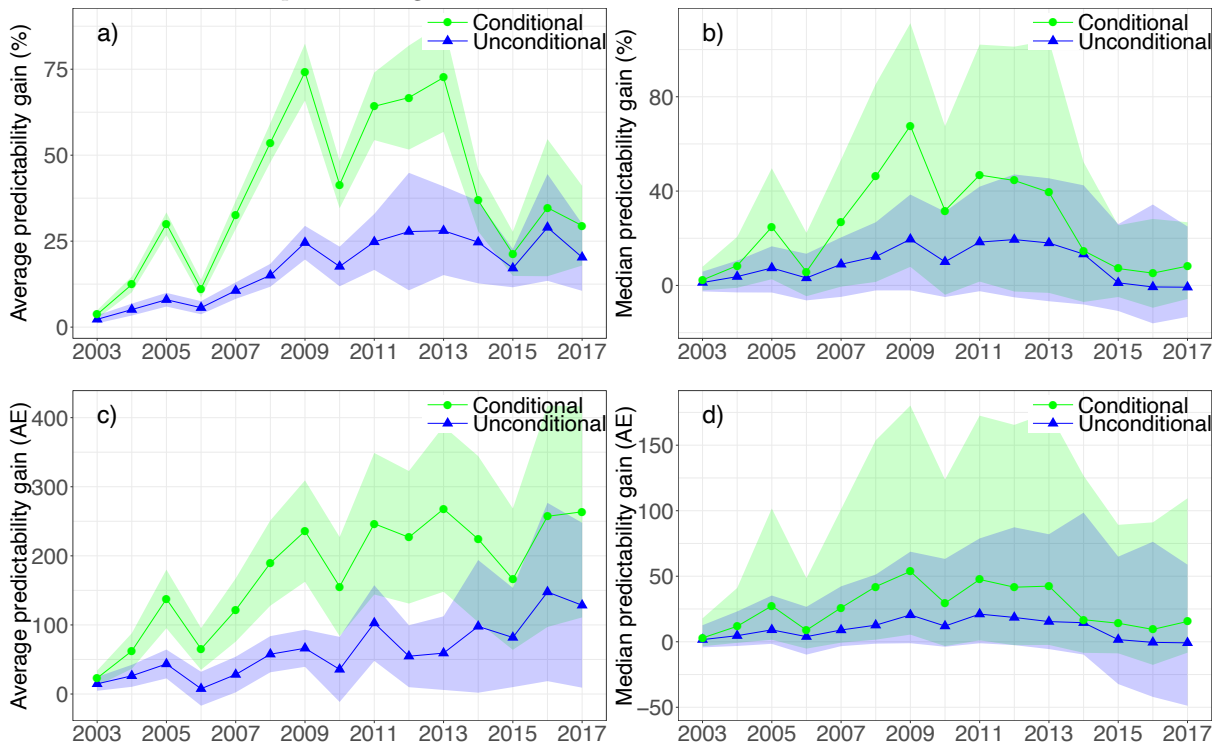


Figure 10: Prediction results based on **IPC 4-digit codes**. a) Average predictability gain based on $PG1$, Eq. (11). Shaded areas indicate twice the standard error. b) Median predictability gain based on $PG1$. Shaded areas indicate the interquartile range. c) Average predictability gain based on $PG2$, Eq. (12). Shaded areas indicate twice the standard error. d) Median predictability gain based on $PG2$. Shaded areas indicate the interquartile range.



References

- J. Alstott, G. Triulzi, B. Yan, and J. Luo. Mapping technology space by normalizing patent networks. *Scientometrics*, 110(1):443–479, 2017. doi: 10.1007/s11192-016-2107-y.
- J. P. Elhorst. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, volume 479. Springer, 2014.
- B. H. Hall, A. Jaffe, and M. Trajtenberg. Market value and patent citations. *RAND Journal of Economics*, pages 16–38, 2005.
- L. Kogan, D. Papanikolaou, A. Seru, and N. Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712, 2017. doi: 10.1093/qje/qjw040.
- F. Lafond and D. Kim. Long-run dynamics of the US patent classification system. *Journal of Evolutionary Economics*, pages 1–34, 2019. doi: 10.1007/s00191-018-0603-3.
- W. Wang and J. Yu. Estimation of spatial panel data models with time varying spatial weights matrices. *Economics Letters*, 128:95–99, 2015. doi: 10.1016/j.econlet.2015.01.021.