A Network-Based Automated Approach for Identifying Technological Spillovers with an Application in Solar Photovoltaics

Background and rationale

Innovation spillovers is a central concept in theories of technological change, but detailed empirical studies have been limited. Literature suggests that breakthrough innovations arise from novel combinations of more technologically "distant" prior inventions. How can we systematically understand the importance of spillovers from more distant domains in the development of a technology?

Utilizing advances in multiple academic disciplines (computer science, econometrics, and network analysis), this paper develops a new methodology to identify the role of spillovers in the historic advancement of a technology domain. We demonstrate our method by studying the development of solar photovoltaic (PV) technology over the period 1901– 2018, quantifying the relative impact of prior art from within the PV sector and from outside PV on the rate and direction of new PV inventions. The results of our work can help inform public policy, such as the organization and management of public and private research laboratories oriented to developing breakthrough, use-driven research. Our demonstration of this methodology in PV has particular importance for climate change mitigation and energy transition.

Methods and Data

This paper develops a network-based automated approach to attributing the technological progress in a field to prior art developed in other fields, so-called "technological spillovers." Our methodology consists of three unsupervised steps, detailed below. First, we utilize **patent citation networks** to build a network of patents within a field. Second, we utilize **natural language processing** to quantitatively measure the "technological distance" between patents. Finally, we use **econometric techniques** to identify patterns in the technological distance between cited and citing patents at points in the citation network that suggest particular significance for the sector. These patterns quantify the dependence of high network importance on citation relationships with more distant prior art.

The first step of our method is to apply network analysis techniques to identify the patents that represent key positions in that sector's technological trajectory. Technological trajectories reflect the main knowledge flows within a patent citation network and provide a holistic view of the shifting technological paradigm over time. Using this approach, we can examine how spillovers affect the changes in the main trajectories over time, rather than the role of spillovers on just a small set of "breakthrough" patents.

Once we identify the main innovation trajectories, we develop a measure of technological distance between patents. To accomplish this, we apply the Latent Dirichlet Allocation (LDA) algorithm, a topic modeling method, to categorize patent abstracts. LDA creates a reduced-dimension representation of documents within a corpus based on word co-occurrence within documents. This approach allows us to calculate technological distance with simple measures, such as Euclidean distance of the frequency of topics within each patent abstract.

Finally, we develop an econometric model to characterize the influence of spillovers on the evolution of innovations along the main paths. Our econometric model gives us a quantitative characterization of how more distant technologies "enter a technology stream" over time. Specifically, the outcome variable ("y variable") is the technological distance between each sequential pair of patents along the main paths. A small distance between two adjacent patents on the main path implies a relatively more incremental improvement in the technology. A large distance between two adjacent patents on the main path signals a broader and more significant shift in the technological paradigm, which we hypothesize is due to the incorporation of more technologically distant prior art. The independent variables of interest ("x variables") relate to the distance between the main path patents and the patents that these patents cite. These x variables reflect the distance (and diversity) of the "knowledge influx" into main-path patents. The further the distance between the main path patents and the cited patents, the greater the main path patents incorporate spillover inventions.

We operationalize our methodology using the PATSTAT for our empirical analysis. Solar PV related patents and their citation networks are identified using the Cooperative Patent Classification (CPC).

Anticipated results

We have fully developed our methodology and will be prepared to present quantitative results at the conference. We plan to present a systematic overview of the evolution of technological trajectories in PV technologies over time. By applying our method to PV patent data, we will display visualizations of the trajectories of PV innovations over time and specific technological breakthroughs and spillovers in the history of PV development.

We also will present a quantitative characterization of the impact of spillovers on innovation. Our econometric models will allow us to estimate the extent to which significant shifts along the main trajectory of solar PV innovation can be attributed to the incorporation (as identified in prior art citations) of technologically distant patents (i.e. spillovers).

Significance

We develop an automated approach, combining network analysis, natural language processing, and econometrics, to quantify the impact of spillovers on technological evolution. The results of applying this method to solar PV demonstrate the contribution of other technology sectors to the development of technology critical for addressing energy and environmental challenges. The results can also help identify possible new roles and types of activities for public policy and research organization design, complementing other insights on the significance of public and private R&D funding and deployment support. This method is fully unsupervised and could be applied to climate-related technologies in other domains. Further, our research group will be combining the fully unsupervised methodology developed here with qualitative interviews with experts and economic modeling of technology cost and performance to assess the validity and implications of this methodology.