NetSciX2026: International School and Conference on Network Science
February 17th -20th, 2026 – Auckland, New Zealand

Bipartite Patent Networks Reveal Hidden Diversity in Innovation: Communities, Market Structure, and Fitness–Complexity

Keywords: Bipartite Networks; Community Detection; Market Structure (HHI/Gini); Fitness–Complexity (ECI/PCI); Patent Analytics

Extended Abstract

Mapping innovation systems through networks is critical for understanding how knowledge creates market value and informs university–industry–government (UIG) policy. Traditional indicators such as citations or licensing are retrospective and delayed, limiting proactive decision-making. By contrast, network-based features available at the pre-grant stage can provide earlier signals for guiding R&D portfolios and policy interventions[1].
Ex-ante assessment of patent value faces structural challenges: patent systems are heterogeneous and nested, with many specialists coexisting alongside a few diversified incumbents. One-mode projections collapse this diversity and obscure early signals[2][3]. To address this, we preserve the two-mode structure by modeling firm–technology and inventor–technology bipartite networks[2][3]. We binarize firm/inventor–technology relationships using Revealed Comparative Advantage (RCA) to mitigate scale bias[4][5] and compute two complementary indicators: a linear measure of diversification and ubiquity (ECI/PCI)[4] and a non-linear fitness–complexity metric[5]. The former highlights the breadth of a technology portfolio, while the latter captures actors' technological capabilities and the difficulty of mastering specific technologies. This dual specification provides complementary perspectives on structural positioning. Market concentration within technology communities is further assessed using HHI and Gini indices.
We prototype the framework on Indian patent data(2010–2011 train/validation, 2019 test; ~125k records). The pipeline includes IPC/CPC subclass parsing, name disambiguation, RCA binarization, bipartite community detection, and indicator computation. Together, these steps constitute a novel patent analytics framework that integrates and extends prior methods [1]. The framework transforms raw data into effective indicators of innovative activity and establishes a reproducible testbed for structural comparison. Predictive modeling is deferred to future large-scale experiments on PATSTAT.
Figure 1 displays a bipartite network for Community 3 (wireless technologies), linking the top-15 firms (left) to IPC subclasses (right). Link widths indicate patent counts, and major hubs are evident, such as Qualcomm. In this network (3,041×50), enterprise Fitness shows a moderate association with high-Complexity coverage (Pearson $r \approx 0.41$; partial $r \approx 0.56$ after controlling for firm size and diversity). Rank-based correlations (Spearman $\rho \approx 0.20$; Kendall $\tau \approx 0.17$) confirm robustness. Firms in the top-10% by Fitness exhibit nearly a tenfold lift in high-Complexity adjacency compared to baseline, with cumulative recall AUC≈0.41, suggesting that Fitness contributes non-random information about structural positioning near complex technologies. By contrast, linear indices (ECI/PCI) tend to prioritize diversified incumbents (e.g., Qualcomm in wireless standards), while the nonlinear Fitness–Complexity framework highlights smaller, specialized firms (e.g., TCS, RIM) and rare, complex subclasses (e.g., B62C). These contrasts point to the complementary perspectives offered by linear and nonlinear approaches rather than a single dominant measure.

Overall, the structural positioning revealed here, particularly the identification of specialized firms with high-Fitness, suggests potential as a leading indicator of future technological breakthroughs and market value. This highlights the promise of our framework, motivating validation on larger global datasets.

Table 1. Top-3 Firms in Community 3 by Fitness (normalized). (2010&2011)

firm	fitness	ECI	patent_count_in_comm
QUALCOMM INC	136.81	0.29	2527
TATA CONSULTANCY SERVICES	98.42	0.41	277
RESEARCH IN MOTION	70.11	0.02	162

Table 2.Representative subclasses by normalized Complexity in Community 3. (2010&2011)

tech	complexity	PCI	edges_unweighted	edges_weighted
B62C	8.37	0.10	1	1
G05F	0.20	0.08	40	54
H03J	4.06	0.11	3	3

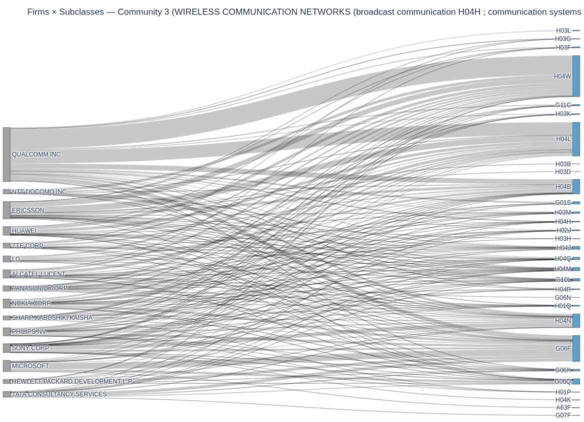

Figure 1. Bipartite Network of Firms × IPC Subclasses in Community 3 (Top-15 Firms)

This study uses only public, aggregate metadata with no sensitive personal data. Disambiguation thresholds and RCA cutoffs were validated across parameter sweeps to prevent overrepresentation of noisy actors. No generative AI was used for content creation; LLMs were only employed for editorial refinement under human oversight, in line with conference policy.

References

[1] Mejía and Y. Kajikawa. Patent research in academic literature: landscape and trends with a focus on patent analytics. Front. Res. Metr. Anal., 2025.
[2] M. Barber. Modularity and community detection in bipartite networks. Phys. Rev. E, 2007.
[3] S. Fortunato and D. Hric. Community detection in networks: A user guide. Phys. Rep., 2016.
[4] C. A. Hidalgo and R. Hausmann. The building blocks of economic complexity. PNAS, 2009.
[5] A. Tacchella et al. A new metric for countries’ fitness and products’ complexity. Sci. Rep., 2012.
…