The n-class model for the US(PTO) patent network

Below we outline the general block strategy for creating an adjacency matrix A for an n-class model. Below, we specify what we believe to be the most important actors: patents, inventors, and firms. We propose to also include other classes such as: technologies, examiners, lawyers, country-inventors.

$$\left[\begin{array}{c|c:c:c} \textbf{PTO} & & & \\ \hline & \textbf{(patents)} & & \\ \hdashline & & \textbf{(inventors)} & \\ \hdashline & & & \textbf{(firms)} \end{array}\right]$$

From this block matrix above, we need to determine (1) how to form the matrix, (2) how to address reducibility (e.g., super-node STATIC), (3) how to block normalize, (4) how to address PROPENSITY/TEMPORAL CONSTRAINTS/SPAMMING BIAS

To be explicit in our understanding of the problem and not have concerns with conformability, we will describe the block matrices as follows:

(PP) Core patent-citation matrix, square, $\mathrm{pp}_{r,c}$, rows represent backward citations, columns represent forward citations.

(IP) Inventor-Patent structure, not square, $\mathrm{ip}_{r,c}$, rows represent inventors, columns represent patents, and the entries represent a truth table of a link structure based on inventor authorship. (By construction, every patent needs at least one inventor, so if the data is missing, what should we do? Possibly create a MISSING node?)

(FP) Firm-Patent structure, not square, $\mathrm{fp}_{r,c}$, rows represent firms, columns represent patents, and the entries represent a truth table of a link structure based on firm assignment. By construction, a patent requires at least one firm, but does not require a firm (the inventor can maintain assignment/ownership). For this reason, we will introduce a NULL node to capture the information that no firm was assigned, but the data is not MISSING (again, possibly create a MISSING node as well?)

We emphasize this nomenclature by updating the matrix:

$$\left[\begin{array}{c|c:c:c} \textbf{PTO} & & & \\ \hline & \textbf{(PP)} & & \\ \hdashline & & \textbf{(II)} & \\ \hdashline & & & \textbf{(FF)} \end{array}\right]$$

From these basic premises, we have options on creating the diagonal block structures:

(PP) Anchored to our traditional Patent Rank model, we have several methods to form the network, include citations, and possibly their strength (combined model as structure + class match). We may want to readdress some of these choices if we include technology as its own class with nodes/edges.

$(II) = \mathbf{0}$ We could just say we have no network information about inventor collaborations.

$(II) = \mathbf{(IP)(PI)}$ We could multiple the inventor-patent network structure by its transpose to identify the collaboration among inventors (i_1 worked with i_3).

$(II) = \mathbf{(IP)(PP)(PI)}$ We could create a quadratic form to weight the associations to include the patent-citation structure. This would capture "influence" but not direct collaboration.

$(II) = \mathbf{(IP)(PI)} + \mathbf{(IP)(PP)(PI)}$ We could create a combined model where we identify the "structure" and "influence".

We could respectively define (FF) following analogous structures, or we could create additional nested structures: patents are created by inventors AND assigned to firms.

$$(FF) = \mathbf{(FP)(PI)(IP)(PF)} \quad ??$$

$$(FF) = \mathbf{(FP)(PI)(PP)(IP)(PF)} \quad ????$$

$$(FF) = \mathbf{(FI)(IP)(PI)(IF)} \quad ??????$$

$$(FF) = \mathbf{(FI)(IP)(PP)(PI)(IF)} \quad ????????$$

Alternatively, we could argue that patents are owned by firms who hire inventors to create them, altering the form of the construction of (II) in a nested manner. We need to think carefully about such constructions both mathematically and conceptually.

Additionally, determining the off-diagonals need to be considered in the construction of the entire matrix.

Once we consider the construction, we then would need to consider the SUPER-NODE. I prefer the single (PTO) supernode at the highest level (the STATIC model) with possible NULL and/or MISSING nodes within specific blocks (but not true "dummy nodes").

In terms of normalization, I also intuitively prefer the row-normalization within a block but there are other possible methods.

A proposed consideration:

$$\left[\begin{array}{c|c:c:c}
\textbf{PTO} & & & \\
\hline
 & (\textbf{patents}) & & \\
\hdashline
 & & (\textbf{inventors}) & \\
\hdashline
 & & & (\textbf{firms})
\end{array}\right] = \left[\begin{array}{c|c:c:c}
\mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\
\hline
\mathbf{1} & (\mathbf{PP}) & (IP) & (FP) \\
\hdashline
\mathbf{1} & (PI) & (\mathbf{II}) = (\mathbf{IP})(\mathbf{PI}) + (\mathbf{IP})(\mathbf{PP})(\mathbf{PI}) & (FI) = (FP)(PI) \\
\hdashline
\mathbf{1} & (PF) & (IF) = (IP)(PF) & (\mathbf{FF}) = (\mathbf{FP})(\mathbf{PI})(\mathbf{II})(\mathbf{IP})(\mathbf{PF})
\end{array}\right]$$