## **Variables :**

pat : Patent Number

fwd: Number of forward citation received by patent

lag\_d: Difference between grant date and application date in years.

noc: Number of claims in the patent. The claims in a patent document delineate the legal scope of the patent in precise language.

gy: year in whic patent was granted

ay: year in which patent was applied at USPTO(United State Patent Office)

noi: number of inventors

umc: 3-digit technology primary classes of patents.

firm id : firm to whom the patent is assigned

MSA: Regional ID : MSA (Metropolitan statistical area) ID for the region of first inventor of patent. Country ID for Non US countries.

tbm: total backward citations made by the patent. Total number of patents cited by this patent

tbru: technical breadth of patent: defined for the focal patent as where refers to the fraction of patents cited by patent i that belong to technology class j. A higher value indicate that patent is broad.

pag: Prior art age : Patents are numbered sequentially, with newer patents being assigned higher patent numbers. Based on this information, we measure the age of the prior art for the focal patent as the average patent number of the patents that it cites (Fleming et al. 2007a). A higher value for this variable indicates that the inventors have built the patent on relatively newer technology.

npr: Non Patent References: the number of non-patent references cited by the focal patent

fpr: Foreign Patent References : number of foreign patent references cited by focal patent.

ft: Firm\_Type: Based On NBER Data types - 7 types of Assigneee: 1 - unassigned; 2 - US non government organization; 3- non US non government organization; 4- US individuals; 5-non US individuals; 6- US government; 7- non US government **Convert this to categories**

cent: Centrality: Teams that occupy more central positions in a network can potentially control both the information flow and the resources of the network. We compute an inventor’s centrality using Freeman's (1979) measure of *betweenness centrality*. A team’s centrality is then calculated by calculating the average *betweenness centrality* across all of the team members (Wasserman and Faust 1994). This measure indicates a team’s ability to absorb (or participate in) information flows and hence, its ability to create valuable innovations check if you can convert this into category

dt : Direct ties : For each patent, we identify the unique inventors who have collaborated with any of the inventors on the focal team during the previous three years but are not current team members. If an inventor has collaborated with more than one team member in the past, the inventor is only counted once. To illustrate, consider a team of two inventors, A and B. Assume that inventor A has ties with C and D at distance 1 and with E and F at distance 2, while inventor B has ties with C and F at distance 1 and with G at distance 2. For this team, the number of unique direct contacts (C, D and F) is 3. We then divide this number by the number of team members to normalize the number of direct contacts for the team (Uzzi and Spiro 2005) . **NOT clear**

it : Number of indirect ties : For each team, we first identify the number of unique inventors that are connected to the inventors in the focal team at a path distance of two or greater, thus excluding the direct ties. The likelihood of information transmission between inventors of the focal team and the external collaborators decreases as the shortest path connecting these increases. To account for this we then utilize the frequency decay measure provided by Burt (1992) in our calculation of indirect ties. The frequency decay function for any external tie at a distance *j* is given as ,where is the number of unique external collaborators of the focal team with a path distance less than or equal to .

rt: We count the number of unique patents on which a pair of inventors of the focal team have worked in the last three years as a measure of repeat ties. (use if else to convert to 1 and 0)

3rd\_Pt: We count the number of unique common inventors with whom a pair of inventors of the focal team (who are collaborating for the first time) have previously worked in the last three years as a measure of third party ties (convert to next positive integer)

rnd\_int: R&D intensity : RD investment /sales - based on data for one year before patent application.

fs: Firm Size : Total no of employess from Compustat - based on data for one year before patent application.

Q1, grpahs between fwd and gy?

Can we drop MSA

Can we divide pag into categories?: plot grpath to get range (barplot) then code using if else and then convert into categories then one hot encoding

Same for cent