

## *Network Science Project Proposal*

### Background

In November 2023 and subsequently in April 2024, the Federal Trade Commission issued warning letters to numerous pharmaceutical manufactures about improper patent listings in the FDA's Orange Book – a publicly available resource on patents and product characteristics of small molecule drugs. It is commonly believed that these improper patent listings serve to deter generic entry by signaling to prospective generic entrants that a pharmaceutical has a stronger legal barrier to entry than it realistically does. For example, an inhaler may improperly list patents in the Orange Book related to mechanical parts or methods of manufacturing in addition to acceptable patent listings on the active ingredient, method of use, or formulation. While this practice first caught the attention of the FTC in the early 2000's, its economic outcomes, which most of my research explores, have remained immensely understudied. Currently, I am working on multiple papers related to this topic, one of which is a descriptive statistics paper in collaboration with a legal expert, which explores the nature of delisted Orange Book patents over time. The idea of that paper is to gain deeper insight into what types of patents were removed from the Orange Book and why, which also sheds light on the extent of improper patent listing over time. For this paper, I was responsible for manually classifying patent claims and collecting continuity data for over 700 delisted patents. This has left me with an extremely rich dataset on a subsection of pharmaceutical patents, which I believe will be ideal for a network analysis.

### Research Questions

My paper seeks to answer several lingering questions that I have related to my Orange Book patent research, and ideally, provide additional insight to my working papers. When reading over each delisted patent, I noticed that many of them have overlapping claims and striking similarities – some patents being nearly indistinguishable in their content despite having different patent numbers. When collecting continuity data, I realized that many of these patents indeed belong to the same family, frequently being continuations or divisions of one another. Using network analysis, it will be possible to uncover the extent of connectivity among removed patents. In particular, I want to answer how many patents in a particular family get delisted when at least one is removed. Additionally, centrality measures on delisted patent nodes will provide an important descriptive layer to my work. Perhaps most important, the number of strongly and weakly connected components will describe how many distinct patent families exist in the delisted patent network.

### Data

To construct my dataset, I drew from several data sources – the first of which was the FDA Orange Book itself. The Orange Book is updated annually (with monthly supplements) with non-accessible data in PDF-book form. To digitally construct this data, I appended 2016 – 2024 data to NBER's digitized Orange Book dataset from 1985 – 2016. This allowed me access

to all patent and exclusivity information ever published in the FDA Orange Book, from which I wrote a script to obtain all patents ever removed prior to their expiration date. For example, if a patent was set to expire in 2030 but was last listed in the 2018 Orange Book edition, this patent would get tagged for my delist dataset. To explore the nature of patents that have been delisted from the Orange Book, I manually classified all delisted patents and their independent claims using Google Patent data. Additionally, using continuity data from Patent Center, I was able to obtain a dataset that will be useful for network analysis. In the continuity data, I have each unique delisted patent connected to parent and children patents, which are also separated by relation type, including continuation, division, and reissue.

### Methods

To construct my network, I will reconstruct my existing continuity dataset into a nodes and edges file. In the nodes file, I plan to tag the delisted patents of interest, which will allow for easier visualization. Additionally, I plan to merge in other patent attributes from my patent classification set, such as claims types and patent categorization (primary, secondary, etc.), to add additional information to the network. In the edges file, I will restructure the data to connect parent patents with children patents, including three different edge types (continuation, division, and reissue). Ultimately, this will allow me to generate a directed graph (showing parent-child relationship) with signed edges to indicate relationship type. The graph generation in and of itself will provide a clear image of the delisted patent network, which I may even include in my related work. In the table below, I used sample data from one delisted patent (#7767223) to generate what the edges file will look like.

Parent	Child	Relationship
7033605	7767223	Continuation
6699493	7767223	Continuation
7767223	7846468	Continuation
7767223	9283178	Division
7767223	8088407	Division
7767223	7625582	Division
7767223	9592242	Division
7767223	8828446	Division
7767223	8043628	Division