Modeling Drug and Medical Device Innovation as Temporal Sequences using EventFlow

(and a few networks)

33rd Annual HCIL Symposium

June 26, 2016
College Park, Maryland
Pennsylvania Innovation Networks 1990 – 2007
Emergence of Philadelphia Biopharma cluster and Pittsburgh Nuclear Cluster
Modeled with Pajek & KING

HCIL Partnership Roots

2010

ME: “It’s cool, but… How do I make it useful?”

BEN: “You must use NodeXL”

2015

ME: “Obiwan Shneiderman, you are my Jedi Master”

BEN: “EventFlow”
How long does innovation take?

From, say, first patent application to new product in the marketplace

Why does it matter?

Why use EventFlow?
Innovation

A process of transforming knowledge and scientific research into a new product in the marketplace.

Think of that process as a sequence of related activities. With this intended outcome.
Innovation

Each activity has inputs, outputs, associated documents and artifacts
With this intended outcome
Each activity involves people and organizations producing *intermediate* outcomes, contributing to this intended innovation outcome.
The people and organizations from each activity create an event network. Activities become event sequences through shared people and organizations, citations, and other linkages.
Data Model Sources

- NIH
  NSF
  StarMetrics
  USA Spending
  University sponsored research data
- Elsevier
  Web of Science
  PubMed
  arXiv
- USPTO
  PatentsView
  ReedTech
  WIPO
  NIH Related Patents
- SBIR / STTR
  USA Spending
  CrunchBase
  Trade publications
  Websites
  Accelerators / Incubators
- CrunchBase
  US FDA
  ClinicalTrials.gov
  TechCrunch
  Trade publications
  Websites
  University licensing
- USPTO
  (Trademarks)
  Thomas.net
  NSNCenter
  UPC Codes
  USA Spending
  US FDA

Linkages

Inputs / Outputs

Activities

Organizations

People

Identity / Relationship

linkages mean that connections must be made based on matching names, dates and other data.
Aggregated event networks: Innovation Ecosystem

The regenerative medicine cluster in Howard County, MD
Innovation Metrics

Some are based on inputs
- Research
- Grant
- Publication

Some are based on outputs
- Invention
- Patent
- Prototype
- Investment
- Beta Product
- Contracts
- Product

Some are comparative indexes
- Commercialization

Activities
- Research
- Invention
- Proof of Concept
- Commercialization

Organizations
- Grantee
- Assignee
- Startup
- Company

People
- PI
- Author
- Inventor
- PI
- Founder
- Entrepreneur

Some are based on talent

Some are based on organizations & resources

None are based on intended outcome
Modeling Innovation Sequences with EventFlow

We use newly developed EventFlow software to model innovation in drugs and medical devices from multiple datasets:

- RePORTER_PATENTS_C_ALL
- RePORTER_CLINICAL_STUDIES_C_ALL
- CTTI AACT Database
- FDA Orange Book (drugs)
- Drugs@FDA
- Pre-Market Approvals (PMA) (med devices)
- SBIR/STTR (pending)
- CrunchBase (pending)
- NSF (pending)

Supporting and core data sources

- NIH RePORTER
- PatentsView
- USA Spending
- STARMETRICS

http://hcil.umd.edu/eventflow/
A Quick Tour of EventFlow

Each product (drug or medical device) is a record in EventFlow (34,331 records)

Event categories:
- Clinical Trials (commercialization activity)
- FDA Approval (proxy for product launch)
- Patents (invention)
- Research
Product-Based Innovation Metrics

Temporal Metrics

How long does innovation take?
How many activities are involved?
What types?
In what sequence?
How long does each take?
Are there gaps?
Is the sequence pattern common or rare?
How long does innovation take? (drugs)

From:
Patent application → FDA approval

(26 products)
How long does innovation take? (drugs)

From:
Patent application → FDA approval (product launch)

(884 drugs in the FDA Orange Book)
How long does innovation take? (med devices)

From start of clinical trial ➔ FDA approval

FDA Approval after Clinical Trial

FDA Approval during Clinical Trial
What’s Next?

- Continued data enhancements
  - Work with FDA and NIH to improve data quality and matching across datasets
  - Add SBIR and other datasets
- Descriptive statistics are important
  - They don’t exist now
  - Grouping drugs and devices by types or properties to improve confidence intervals
- Sequence characteristics
  - How many events? What types? What order?
  - Gaps and overlaps
- Exploration of Data
- Hypothesis Testing
What’s Next?

• Development of new metrics for Science, Technology and Innovation
• Testing hypotheses about performance of spatial – social systems
Broader applications of temporal metrics: the Illinois Battery Cluster

Innovation Ecosystems

- research component
- Industry component
- Bridging component
The Innovation Ecosystem and the Valley of Death

A network representation of the valley of death

Basic Research
Agencies & Institutions
Applied Research
Research Component
Invention
IP & Startups
Proof-of-Concept
Sponsored Research
SBIR / STTR
Entrepreneurs & Investors
Venture Stage
Commercialization (Industry Component)
Angel Stage
Seed Stage
Time
Emerging Theory & Research

What’s in the Bridge?

- **Working Hypothesis**

- Regions with denser, more connected bridging components will be characterized by **faster innovation sequences** and more innovation sequences leading to new products.

Measured using new temporal metrics
Thank you to the US Economic Development Administration and the National Science Foundation for supporting this research