Does Death of Distance Hypothesis hold true for knowledge networks?

International Patent Collaboration and ICT Diffusion

Esha Banerjee
Advisor: Dr. Richard W. Evans

Death of distance hypothesis

Information and communication technologies (ICT) will make many things that were previously limited by distance, primarily communication and transmission of information, no longer affected by distance.

Geography & Knowledge networks

Geographic constraints and national borders impede the diffusion of knowledge (evidenced in R&D collaborations).

Evolution of Knowledge networks

Do such constraints persist for the networks with the diffusion of ICT over time?

Or,

Has ICT lead to collaborations between countries that did not previously collaborate or has it caused the clusters of collaboration to expand?

Evolution of Knowledge networks

Do such constraints persist for the networks with the diffusion of ICT over time?

Or,

Has ICT lead to collaborations between countries that did not previously collaborate or has it caused the collaborative distances to expand?

Knowledge networks: Co-inventorship

Cross-country links measured through collaborations between inventors of different country pairs (1990-2010)

Data: European Patent Organization, PATSTAT database

ICT diffusion indicators

- Percentage of individuals using the internet 1990-2010
- Percentage of individuals having mobile-telephone subscriptions 1990-2010

Data: World Telecommunication/ICT Indicators Database

ICT diffusion indicators

- Percentage of individuals using the internet 1990-2010
- Percentage of individuals having mobile-telephone subscriptions 1990-2010

Data: World Telecommunication/ICT Indicators Database

Model?

- Gravity model, frequently employed in trade
- Recently used in knowledge networks

Gravity Model

General Form: $F_{i,j} = G.\frac{MiMj}{Di,j}$

Time-variant Indicators

- Population
- Patents filed at country level
- GDP

Data: World Intellectual Property Organization,
World Development Indicators

Time-invariant Indicators

- Distance
- Contiguity
- Time difference

Data: CEPII Database

Time-invariant Indicators

- Common Language

- Colonial Ties

Data: CEPII Database

Estimation

- Pseudo Poisson Maximum Likelihood estimation

- Works well in presence of lots of zeroes in the data

Empirical Analysis

- Increase in number of collaborative patents

- Increase in number of country pairs that are collaborating

Empirical Analysis

- Countries that frequently collaborate have remained largely unchanged over the years

US-CN

US-DE

US-GB

US-JP

US-FR

US-NL

US-CA

CN-DE

Estimation Results

- A unit change in ln(distance) causes collaboration to decrease by ~65%

- A unit change in ln(perc_internet) causes an increase of patent collaboration by 4%

Estimation Results

- ICT indicators are not statistically significant, however it indicates an increase in co-invention when the countries have similar internet diffusion.
- Measures robust across different models
- Language barriers and time-difference remain statistically significant across different models

Estimation Results

- ICT indicators are not statistically significant, however it indicates an increase in co-invention when the countries have similar internet diffusion.

- Measures robust across different models

- Language barriers and time-difference remain statistically significant across different models

Conclusions & Limitations

- No conclusive evidence yet that improving ICTs close the gap of geographical distance when it comes to collaboration at an inventor level across countries

- Missingness, ICT diffusion data quality

Speaker Transition

Topic-wise Analysis on the Federal Deposit Insurance Corporation (FDIC) Enforcement Decisions and Orders

Alice Chung
Computational Social Science

Background

- To secure safe and sound banking practices, federal regulators monitors the banking industry.
  - Federal banking regulators:
    - the Federal Deposit Insurance Corporation(FDIC), Federal Reserve System(FRS), the Office of the Comptroller of the Currency(OCC) and the National Credit Union Administration(NCUA)
  - Issuance of Enforcement Actions

Background

- To secure safe and sound banking practices, federal regulators monitors the banking industry.
  - Federal banking regulators:
    - the Federal Deposit Insurance Corporation(FDIC), Federal Reserve System(FRS), the Office of the Comptroller of the Currency(OCC) and the National Credit Union Administration(NCUA)
  - Issuance of Enforcement Actions

- FDIC issued majority of Enforcement Actions.
  - Publicly published the full text of the formal enforcement actions every month.
  - Contain consent decrees and opinions.

Purpose

- Exploring the meaningful information
  - Year, state, the names of parties, the reasons for the enforcement actions.
- In-depth, what can I get from the text of enforcement actions?
  - Focus on the reasons the action types to classify the documents
  - Clustering by topics, and study the shift trend

Research question(s)

- In topic-wise, how do recent trends in enforcement actions in the banking industry and how it varies through the time?
- What factors or reasons cause the issuance of enforcement actions?

Literature Review

- Previous text analysis on accounting:
  - Mostly on annual reports, chairman's statement, and financial reviews and other accounting related statements or communications (Gary et al, 1994, Macolm and Richard, 2000, and Fouad, 2000)
- The effects of bank enforcements on the bank's performance
  - The effects of Bank enforcements on the terms of lending (Deli et al, 2016)
  - The effects of Bank examinations and enforcements actions on the behavior of problem banks (Curry et al, 1999)
- Examination on the proper role of the discretionary enforcement in banking capital regulation
  - Capital regulation system that relies heavily on individual bank (Hill, 2012)

Literature Review

- Previous text analysis on accounting:
  - Mostly on annual reports, chairman's statement, and financial reviews and other accounting related statements or communications (Gary et al, 1994, Macolm and Richard, 2000, and Fouad, 2000)
- The effects of bank enforcements on the bank's performance
  - The effects of Bank enforcements on the terms of lending (Deli et al, 2016)
  - The effects of Bank examinations and enforcements actions on the behavior of problem banks (Curry et al, 1999)
- Examination on the proper role of the discretionary enforcement in banking capital regulation
  - Capital regulation system that relies heavily on individual bank (Hill, 2012)

Data

- The text files downloaded from FDIC Enforcement Decisions and Orders database from the period 1990 to 2017
- 9,982 enforcement decisions and orders text documents
- Preprocessing and sample selections (still in progress)
  - Format change : HTML to PDF from 2006
  - Using OCR (Optical character recognition) to read PDF
  - Dirty and Messy
  - OpenRefine
  - Dropping all the documents that can not be accessed using attachment link and read by OCR -> 9,618

Data

- The text files downloaded from FDIC Enforcement Decisions and Orders database from the period 1990 to 2017
- 9,982 enforcement decisions and orders text documents
- Preprocessing and sample selections (still in progress)
  - Format change : HTML to PDF from 2006
  - Using OCR (Optical character recognition) to read PDF
  - Dirty and Messy
  - OpenRefine
  - Dropping all the documents that can not be accessed using attachment link and read by OCR -> 9,618

Data

- The text files downloaded from FDIC Enforcement Decisions and Orders database from the period 1990 to 2017
- 9,982 enforcement decisions and orders text documents
- Preprocessing and sample selections (still in progress)
  - Format change : HTML to PDF from 2006
  - Using OCR (Optical character recognition) to read PDF
  - Dirty and Messy
  - OpenRefine
  - Dropping all the documents that can not be accessed using attachment link and read by OCR -> 9,618

Data

FDIC Enforcement Decisions and Orders

{{04-30-06 p.12540.1}}

[¶12,540] In the Matter of Salt Lick Deposit Bank, Salt Lick, Kentucky, Docket No. 05-220k (2-24-06).

Respondent agrees to pay civil money penalty assessed by the FDIC in the amount of $8,100.

In the Matter of
SALT LICK DEPOSIT BANK
SALT LICK, KENTUCKY
(Insured State Nonmember Bank)
ORDER TO PAY

FDIC-05-220k
{{04-30-06 p.12541.1}}

Salt Lick Deposit Bank, Salt Lick, Kentucky ("Respondent") and a representative of the Legal Division of the Federal Deposit Insurance Corporation ("FDIC") executed a STIPULATION AND CONSENT TO THE ISSUANCE OF AN ORDER TO PAY ("CONSENT AGREEMENT") dated December 22, 2005, whereby Respondent, solely for the purpose of this proceeding and without admitting or denying any violation of law for which a civil money penalty may be assessed, consented and agreed to pay a civil money penalty in the amount specified below to the Treasury of the United States.

After taking into account the CONSENT AGREEMENT, the appropriateness of the penalty with respect to the financial resources and good faith of the Respondent, the gravity of the violations by the Respondent, the history of previous violations by the Respondent, and such other matters as justice may require, the FDIC accepts the CONSENT AGREEMENT and issues the following:

ORDER TO PAY

IT IS HEREBY ORDERED, that by reason of the violations set forth in paragraph 3 of the CONSENT AGREEMENT, a penalty of $8,100 be, and hereby is, assessed against the Respondent. The Respondent shall pay the civil money penalty to the Treasury of the United States.

This ORDER TO PAY shall be effective upon issuance.

Pursuant to delegated authority.

Dated at Washington, D.C., this 24th day of February, 2006.

FEDERAL DEPOSIT INSURANCE CORPORATION

WASHINGTON, D.C.

In the Matter of Peter D. Rohn, Application for Consent to Participate in the Conduct of the Affairs of Any Insured Depository Institution	ORDER GRANTING PERMISSION TO FILE APPLICATION AND APPROVING APPLICATION FOR CONSENT TO PARTICIPATE IN THE CONDUCT OF THE AFFAIRS OF ANY INSURED DEPOSITORY INSTITUTION FDIC-17-0204L

The Federal Deposit Insurance Corporation (FDIC) has fully considered all the facts and information relating to the application filed pursuant to Section 19 of the Federal Deposit Insurance Act, 12 U.S.C. § 1829 (Section 19), by Peter D. Rohn (Applicant), individually, for a waiver of the requirement that an insured depository institution file this Section 19 application and for consent to participate directly or indirectly in the conduct of the affairs of any insured depository institution. The FDIC determines that Applicant's request for a waiver should be granted and that Applicant's Section 19 application to participate in the conduct of the affairs of any insured depository institution should be approved based upon the following:

1. On May 13, 1974, Applicant entered a plea of guilty to three counts of distribution of marijuana in violation of Virginia Code Ann. §18.2-248.

2. The FDIC notes that 43 years have elapsed since the offenses, and the Applicant has had no further convictions or program entries subject to Section 19.

3. The FDIC finds that Applicant has demonstrated satisfactory evidence of rehabilitation.

4. The FDIC finds that Applicant has demonstrated substantial good cause for waiving the requirement that an insured depository institution file a Section 19 application on his behalf and further finds that this requirement should be waived.

5. The FDIC finds that Applicant's participation, directly or indirectly, in the conduct of the affairs of any insured depository institution, in any position, would not pose a threat to the safety or soundness of any insured depository institution or the interests of depositors, nor would such participation threaten to impair public confidence in any insured depository institution.

ORDER

Accordingly, it is ORDERED that Peter D. Rohn is permitted to file his Section 19 application on his own behalf and that this Section 19 application to participate directly or indirectly in the conduct of the affairs of any insured depository institution is APPROVED, provided that prior to participating in the affairs of an insured depository institution, Applicant must supply that insured depository institution with a copy of this ORDER GRANTING PERMISSION TO FILE APPLICATION AND APPROVING APPLICATION FOR CONSENT TO PARTICIPATE IN THE CONDUCT OF THE AFFAIRS OF ANY INSURED DEPOSITORY INSTITUTION, and provided that Applicant must be covered by a fidelity bond to the same extent as others with a similar position at that insured depository institution.

This consent applies only to the offenses described in paragraph 1 above.

Dated this 7th day of September, 2017.

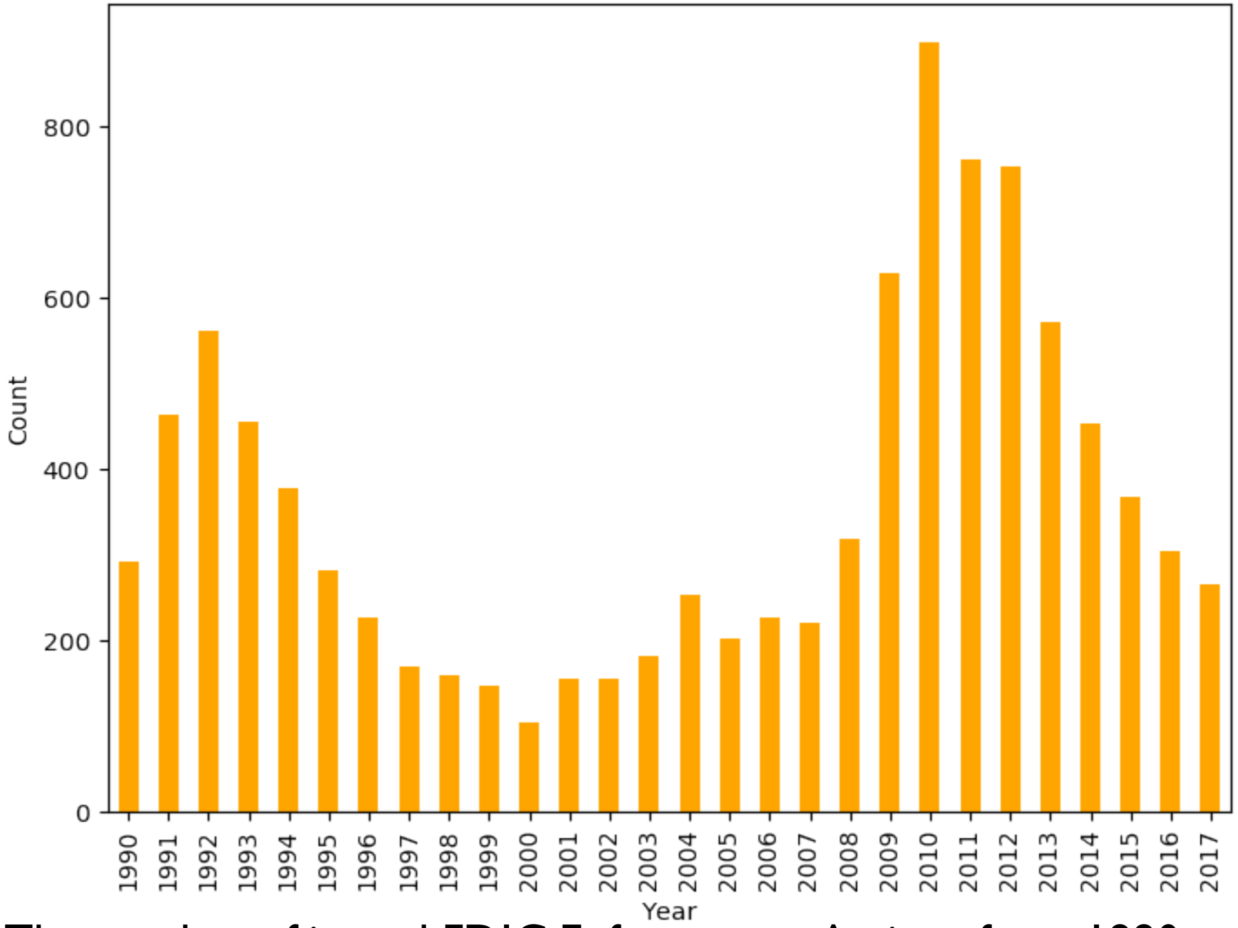
/s/
Patricia A. Colohan
Associate Director
Division of Risk Management Supervision

2

Methodology

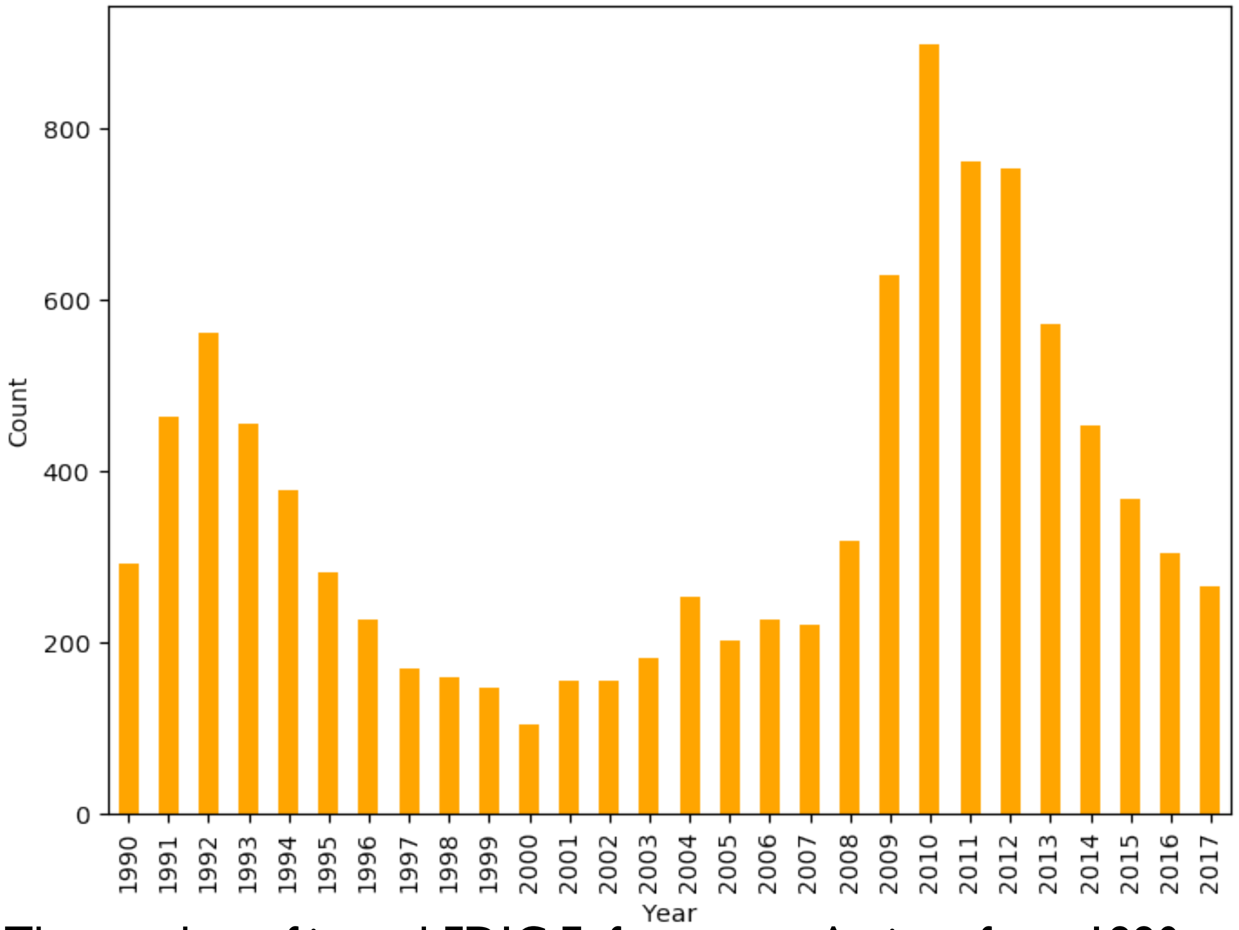
- Text analysis and Topic modeling
  - Latent Dirichlet allocation (LDA) to discovering topics
  - Unsupervised Bayesian machine learning model to detect topics from the large corpora
  - To understand and study these topics to answer the recent trends and variations in enforcement actions in the banking industry

Preliminary Results



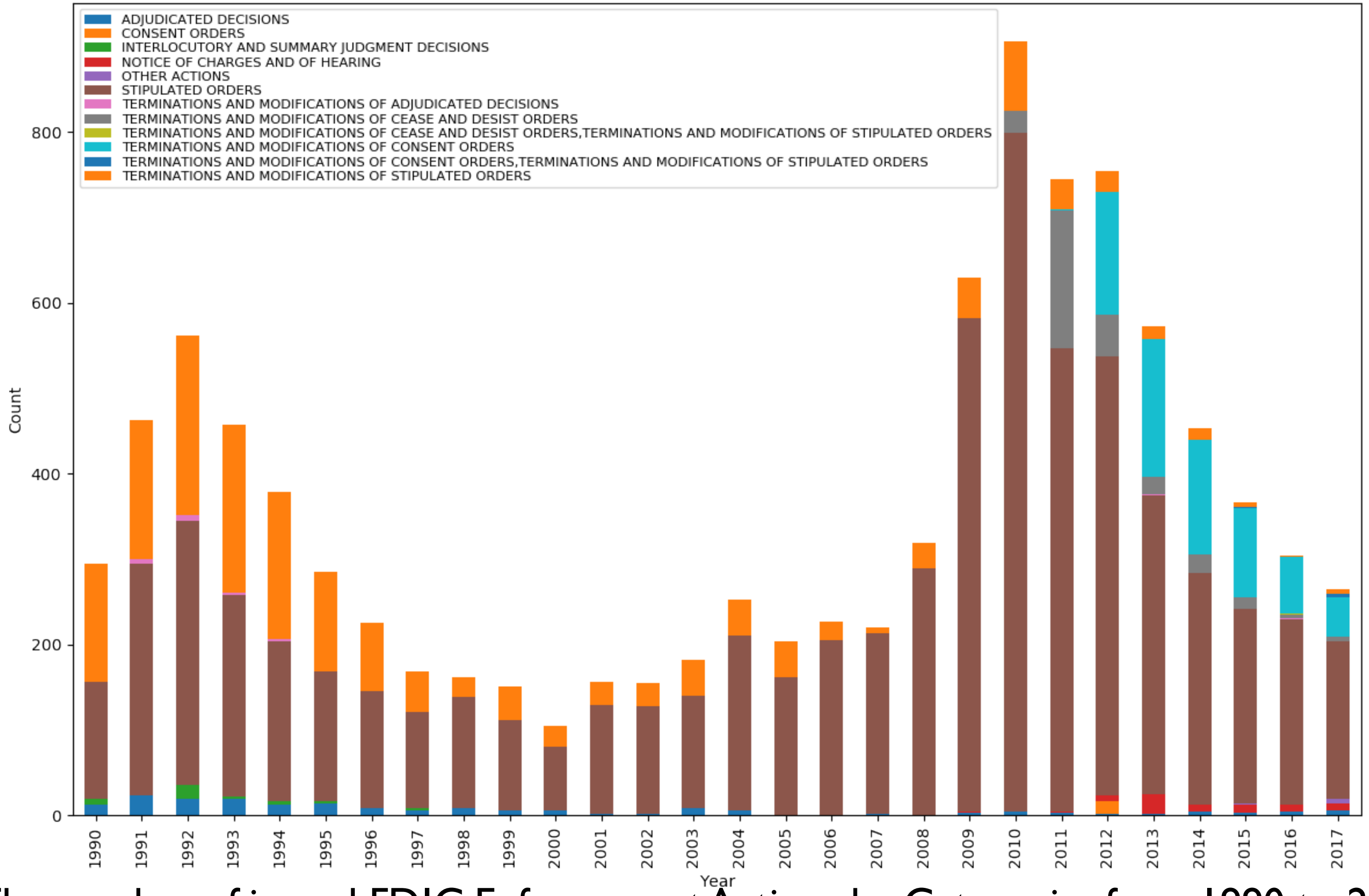
The number of issued FDIC Enforcement Actions from 1990 to 2017

Preliminary Results



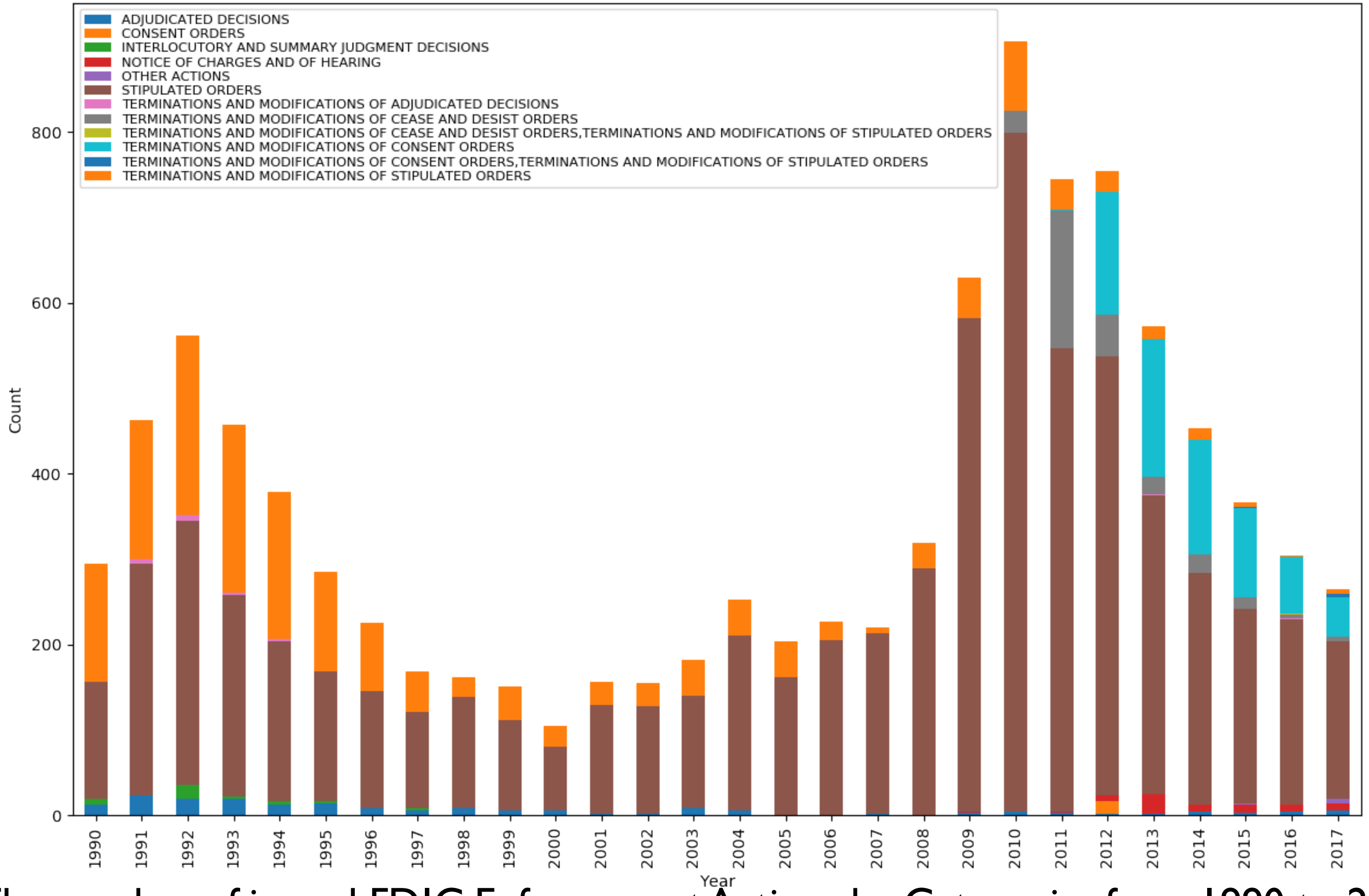
The number of issued FDIC Enforcement Actions from 1990 to 2017

Preliminary Results



The number of issued FDIC Enforcement Actions by Categories from 1990 to 2017

Preliminary Results



The number of issued FDIC Enforcement Actions by Categories from 1990 to 2017

Preliminary Results with LDA(k=45)

Topic	Unigrams
1	'institut', 'notic', "'s", 'liabl', 'insur', 'u.s.c', 'loss', 'pursuant', 'deposit', 'hear', 'law', 'state', 'asset', 'liabil', 'provis', 'act', 'capit', 'director', 'feder', 'ani', 'depositori', 'c.f.r', 'financi', 'amount', 'immedi', 'propos', 'total', 'c', 'find', 'incur', 'day', 'proceed', 'request', 'reason', 'corpor', 'e', 'advanta', 'matter', 'pay', 'practic', 'time', 'b', 'board', 'washington', 'estim’
2	“complianc', "'s", 'board', 'effect', 'manag', 'audit', 'report', 'procedur', 'day', 'director', 'within', 'ensur', 'law', 'system', 'offic', 'implement', 'includ', 'program', 'review', 'ani', 'consum', 'correct', 'polici', 'requir', 'train', 'action', 'b', 'provid', 'region', 'committe', 'provis', 'respons', 'monitor', 'written', 'ii', 'examin', 'develop', 'perform', 'follow', 'regul', 'employe', 'relat', 'minimum', 'document', 'violat’
3	'file', "'s", 'e', 'c', 'respond', 'violat', 'mr.', 'abc', 'act', 'section', 'law', 'fail', 'o~', 'deposit', 't~', 'ani', 'cash', 'f', 'charg', 'u', 'ofth', 'day', 'b', 'insur', 'practic', 'iii', 'forth', 'r', 'virginia', 's.c.', 'caus', 'west', 'set', 'tl~e', 'rosoff', 'account', 'issu', 'washington', 'offic', 'pertin', 'activ', 'serv', '~o', 'defin', 'feder’
4	'ani', "'s", 'director', 'loan', 'effect', 'plan', 'day', 'board', 'includ', 'region', 'within', 'secur', 'capit', 'implement', 'determin', 'manag', 'asset', 'polici', 'requir', 'review', 'report', 'and/or', 'provis', 'materi', 'prior', 'credit', 'b', 'written', 'offic', 'paragraph', 'condit', 'addit', 'oper', 'subsequ', 'commission', 'classifi', 'visit', 'part', 'examin', 'adopt', 'regul', 'collect', 'loss', 'financi', 'develop’
5	'payment', 'oklahoma', 'reimburs', 'processor', 'affili', 'third', 'ach', 'parti', 'entiti', 'reserv', 'tit', 'okla.', 'et', 'stat', 'state', 'seq', 'paid', 'citi', 'third-parti', 'forens', 'dalla', 'account', 'januari', 'balanc', 'seek', 'six', 'octob', 'tulsa', 'regist', 'fifti', 'amount', 'charge-back', 'cash', 'thirty-f', 'thompson', 'event', 'spiritbank', 'mick', 'transit', 'thorough', 'pim', '2order', 'befor', 'd-1', 'fraud’

Preliminary Results with LDA(k=45)

Topic	Unigrams
6	'ani', 'e', 'respond', 'breach', 'practic', 'and/or', 'duti', 'fiduciari', 'particip', 'vote', 'act', 'violat', 'financi', 'institut', 'u.s.c', 'consent', 'section', 'prohibit', 'parti', 'unsaf', 'agenc', 'unsound', "'s", 'enumer', 'institution-affili', 'agreement', 'issuanc', 'determin', 'demonstr', 'conduct', 'feder', 'person', 'b', 'director', 'organ', 'affair', 'serv', 'reason', 'provis', 'transfer', 'c', 'approv', 'effect', 'attempt', 'right’
7	'custom', "'s", 'account', 'august', 'robert', 'check', 'larsson', 'balanc', 'overdraft', 'act', 'new', 'deposit', 'section', 'approxim', 'assess', 'uncollect', 'kite', 'pursuant', 'u.s.c', 'report', 'practic', 'loss', 'activ', 'penalti', 'pay', 'hear', 'e', 'may', 'loan', 'and/or', 'offic', 'total', 'increas', 'amount', 'civil', 'recommend', 'insur', 'despit', 'money', 'notic', 'day', 'jersey', 'februari', 'daili', 'overdrawn’
8	'respond', 'notic', 'loan', 'u.s.c', 'act', 'section', 'e', 'practic', "'s", 'law', 'hear', 'proceed', 'c.f.r', 'assess', 'state', 'pursuant', 'time', 'file', 'financi', 'account', 'violat', 'rule', 'insur', 'prohibit', 'feder', 'unsound', 'amount', 'deposit', 'day', 'charg', 'check', 'pay', 'unsaf', 'offic', 'person', 'corpor', 'penalti', 'fact', 'use', 'breach', 'term', 'request', 'institut', 'answer', 'pertin’
9	'direct', 'section', 'ani', "'s", 'capit', 'act', 'u.s.c', 'c.f.r', 'action', 'b', 'rule', 'pursuant', 'e', 'effect', '1831o', 'region', 'correct', 'provis', 'prompt', 'requir', 'regul', 'director', 'offic', 'insur', 'deposit', 'supervisori', 'plan', 'dure', 'day', 'secur', 'restor', 'feder', 'wherea', 'prior', 'issu', 'materi', 'necessari', 'condit', 'period', 'notic', 'institut', 'appeal', 'written', 'offer', 'practic’
10	“loan', "'s", 'respond', 'nomine', 'boruff', 'number', 'document', 'process', 'rd', 'approv', 'obtain', 'prepar', 'william', 'duti', 'e', 'caus', 'cr', 'unsaf', 'strode', 'practic', 'made', 'powel', 'act', 'lend', 'second', 'unsound', 'proceed', 'offic', 'otherwis', 'breach', 'lexus', 'fiduciari', 'financi', 'use', 'purpos', 'first', 'ani', 'legal', 'signatur', 'execut', 'time', 'borrow', 'sound', 'receiv', 'secur’

Further steps

- More precise preprocessing on the documents
- Speed up problem of running LDA
- Select optimal number of topic

References

- Curry, T.J., O'Keefe, J.P., (1999), "Financially distressed banks: how effective are enforcement actions in the supervision process?", FDIC Banking Rev. 12 (2), 1–18.
- Deli, Yota and Delis, Manthos D. and Hasan, Iftekhar and Liu, Liuling, (2016), "Bank Enforcement Actions and the Terms of Lending", Gabelli School of Business, Fordham University Research Paper No. 2786892.
- Flannery, M.J., 1998. "Using market information in prudential bank supervision: A review of the U.S. Empirical Evidence." Journal of Money, Credit and Banking 30, 273–305.
- Fouad K. Alnajjar, (2000), "Determinants of social responsibility disclosures of U.S. Fortune 500 firms: An application of content analysis", Advances in Environmental Accounting & Management (Advances in Environmental Accounting & Management, Volume 1) EmeraldGroup Publishing Limited, pp.163 – 200.
- Frazier, K., Ingram, R., & Tennyson, B, (1984), "A Methodology for the Analysis of Narrative Accounting Disclosures", Journal of Accounting Research, 22(1), 318-331.
- Hill, Julie A. (2012), "Bank Capital Regulation by Enforcement: An Empirical Study,", Indiana Law Journal: Vol. 87 : Iss. 2 , Article 4. Lambert, Thomas, (2017), "Lobbying on Regulatory Enforcement Actions: Evidence from U.S. Commercial and Savings Banks", Management Science, Forthcoming.
- Malcolm Smith, Richard J. Taffler, (2000), "The chairman's statement - A content analysis of discretionary narrative disclosures", Accounting, Auditing & Accountability Journal, Vol. 13 Issue: 5, pp.624-647.
- Previts, Gary John, Robert J. Bricker, Thomas R. Robinson, and Stephen J. Young.,(1994), "A content analysis of sell-side financial analyst company reports", Accounting Horizons 8, (2) (06):55.
- Roman, Raluca A., (2017), "Collateral Benefits? Impact of Supervisory Enforcement Actions against Banks on Their Corporate Borrowers",Available at SSRN: https://ssrn.com/abstract=2919009 or http://dx.doi.org/10.2139/ssrn.2919009

Speaker Transition

Neighbourhood Disorder and Teen Pregnancy Rates

An Application of the Broken Windows Theory

Chelsea Ernhofer

Research Question

Can the Broken Windows theory be used to help explain and predict teen birth rates in New York City?

Past Literature // Teen Pregnancy and Birth

- Individual
    - Use of condoms
    - Individual sexual experience
- Educational
    - Comprehensive sexual education vs. abstinence education
    - Contraceptive distribution
- Social
    - Social capital and social cohesion
    - Social networks and role models
    - Crime rates
- Ecological
    - Built Environment

Past Literature // Teen Pregnancy and Birth

- Individual
  - Use of condoms
  - Individual sexual experience
- Educational
  - Comprehensive sexual education vs. abstinence education
  - Contraceptive distribution
- Social
  - Social capital and social cohesion
  - Social networks and role models
  - Crime rates
- Ecological
  - Built Environment

Past Literature // Broken Windows Theory

- Physical environment communicates levels of disorder
- A neighbourhood with increased physical disorder (ie. broken windows) is more likely to experience high crime rates
- Broken Windows theory has been linked to
  - Homicide crime rates
  - Perception of safety/perception of crime
  - Community health

Past Literature // Broken Windows Theory

- Physical environment communicates levels of disorder
- A neighbourhood with increased physical disorder (ie. broken windows) is more likely to experience high crime rates
- Broken Windows theory has been linked to
  - Homicide crime rates
  - Perception of safety/perception of crime
  - Community health

Hypothesized Model

Teen Birth Rate $= \beta_0 + \beta_1$*(disorder)* $+ \beta_2$*(crime rates)* $+$
$\boldsymbol{\beta}$*(demographic controls)*

Capturing Neighbourhood Disorder

StreetScore Data Set

- StreetScore: algorithm designed to apply a score to a street view based on its supposed safety
- Used in past literature to represent neighbourhood disorder rather than safety

Capturing Neighbourhood Disorder

StreetScore Data Set

- StreetScore: algorithm designed to apply a score to a street view based on its supposed safety
- Used in past literature to represent neighbourhood disorder rather than safety

Data // Sources

Neighbourhood Disorder - StreetScore Data // NYC Spring 2014

Teen Pregnancy Rate - New York Department of Health // 2012-2014

Crime - NYC Open Data // 2014

Demographic Controls - American Community Survey // 2014

Data // Cleaning

- Analysis at the zip code level
  - Performed geospatial overlays to sort point data (latitude, longitude) into zip codes
- Removal of outliers
  - Extreme outliers of StreetScore data points were removed
- Normalization of basic counts
  - All demographic variables and crime occurrences were normalized by population within zip code

Table 1. Summary of Model Variables

	N	*Mean*	*St. Dev.*	*Min*	*Max*
Teen Birth Rate	149	17.716	10.803	0.6	44
Order	149	25.158	1.488	19.550	27.821
Crime Rate	149	0.059	0.040	0.016	0.320
Race					
White	149	0.450	0.264	0.019	0.942
Black	149	0.235	0.258	0.000	0.9221
American Indian/Alaskan	149	0.003	0.003	0.000	0.018
Asian	149	0.134	0.134	0.003	0.726
Hawaiian/Pacific Islander	149	0.000	0.000	0.000	0.005
Median Income	149	57874.15	22413.27	19536	115604
Employment Rate	149	0.518	0.076	0.379	0.753
Receives Public Assistance	149	0.015	0.009	0.002	0.043

Methods

Linear Regression // Decision Trees // Random Forests

- Interested in both inference and prediction

- Exploration of linear vs. non-linear modeling

Results

Results // Linear Regression

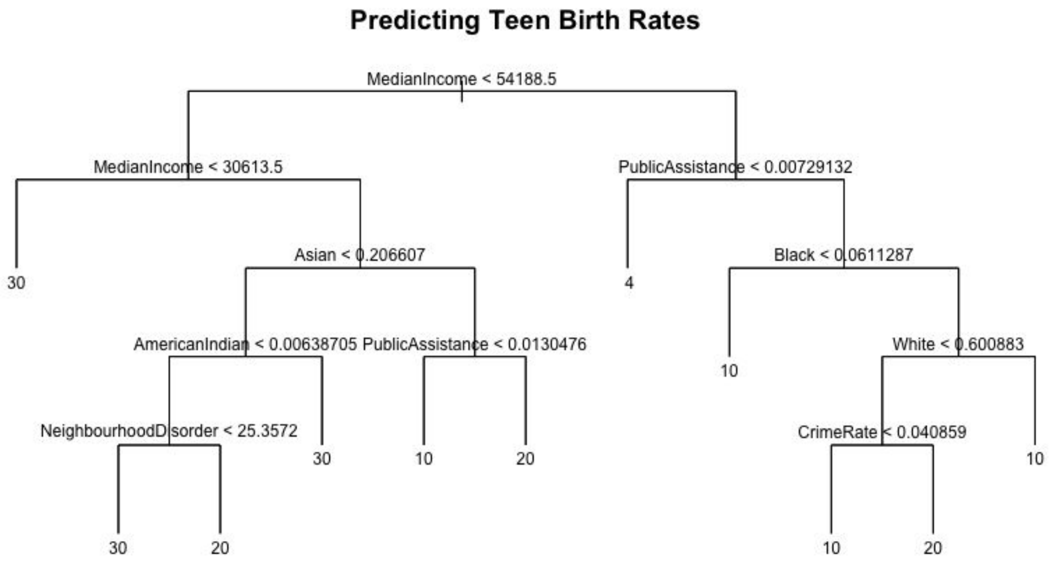
Table 2. Linear Regression Results

	Model
Order	-1.2101360 ***
Crime Rate	0.0003555
Race	
White	-0.0001695 **
Black	-0.0002210 ***
American Indian/Alaskan	0.0023128
Asian	-0.0004194 ***
Hawaiian/Pacific Islander	-0.0056982
Median Income	-0.0002415 ***
Employment Rate	0.0008310 ***
Receives Public Assistance	0.0039694 **

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

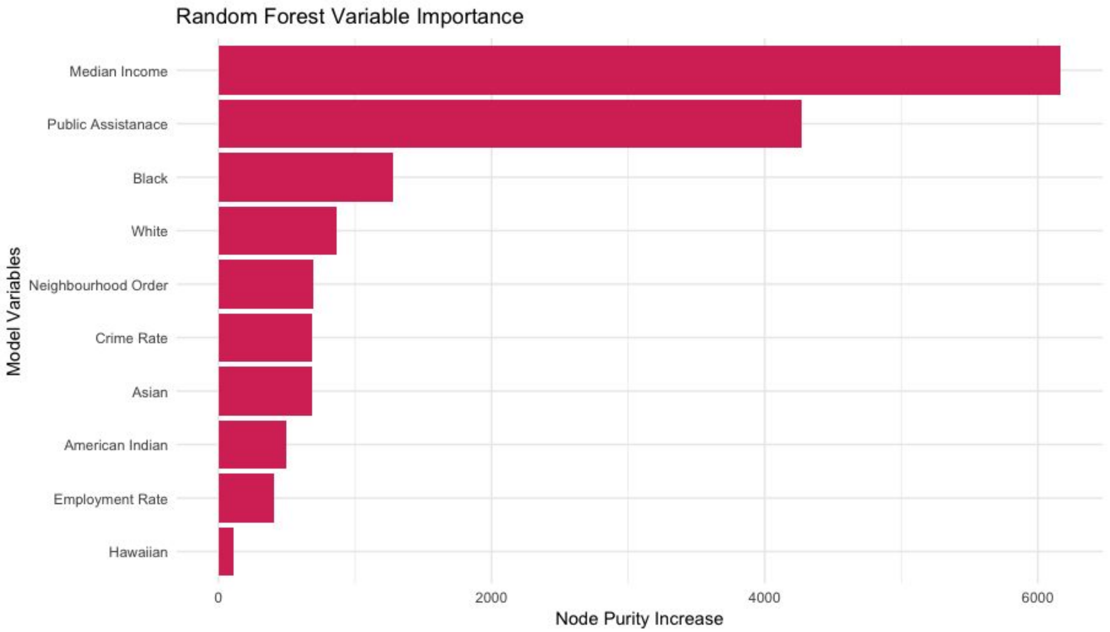
- Adj. $R^2 = 0.777$
- As neighbourhood order increases, teenage pregnancy rates decrease.
- Controlling for crime and basic neighbourhood demographics

Results // Regression Tree



- Demographic control variables such as median income, public assistance, and race appear to be more useful in partitioning than disorder or crime

Results // Random Forest



- Median Income and Public Assistance are the most important indicators of Teen birth rates

Results // Predictive Power

Method	RMSE
Linear Regression	6.592
Regression Tree	4.560
Random Forest	5.330

References

Cohen, Deborah, et al. "Broken windows" and the risk of gonorrhea." American journal of public health 90.2 (2000): 230.
Cubbin, Catherine, et al. "Neighborhood context and sexual behaviors among adolescents: Findings from the National Longitudinal Study of Adolescent Health." Perspectives on sexual and reproductive health 37.3 (2005): 125-134.
Cummins, Steven, et al. "Neighbourhood environment and its association with self rated health: evidence from Scotland and England." Journal of Epidemiology & Community Health 59.3 (2005): 207-213.
Darroch, Jacqueline E., Susheela Singh, and Jennifer J. Frost. "Differences in teenage pregnancy rates among five developed countries: the roles of sexual activity and contraceptive use." Family planning perspectives (2001): 244-281.
Elliott DS, et al. The Effects of Neighborhood Disadvantage on Adolescent Development. Journal of Research in Crime and Delinquency. 1996;33(4):389–426.
Farkas, Brenlea, et al. "The Associations between Neighbourhood Built Characteristics and Physical Activity among Adults: A Systematic Review of Canadian Evidence (breakout presentation)." Journal of Transport & Health 7 (2017): S32.
Leventhal T, Brooks-Gunn J. The Neighborhoods They Live in. Psychological Bulletin. 2000;126(2):309–337
Maimon D, Browning CR. Unstructured Socializing, Collective Efficacy, and Violent Behavior Among Urban Youth. Criminology. 2010;48(2):443–474.
Naik, Nikhil, et al. "Streetscore-predicting the perceived safety of one million streetscapes." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014.
Santelli, John S., et al. "Can changes in sexual behaviors among high school students explain the decline in teen pregnancy rates in the 1990s?." Journal of Adolescent Health35.2 (2004): 80-90.
Stanger-Hall, Kathrin F., and David W. Hall. "Abstinence-only education and teen pregnancy rates: why we need comprehensive sex education in the US." PLoS One 6.10 (2011): e24658.
Weich, Scott, et al. "Mental health and the built environment: Cross–sectional survey of individual and contextual risk factors for depression." The British Journal of Psychiatry 180.5 (2002): 428-433.

Speaker Transition

Predicting Infant Mortality: Minimizing False Negatives

Sushmita V Gopalan

Context

- 32 out of 1000 babies born in India die within a year of their birth (2015)
- Stagnation, despite multiple government schemes
- Tesfaye (2017) in Ethiopia

Context

- 32 out of 1000 babies born in India die within a year of their birth (2015)

- Stagnation, despite multiple government schemes

- Tesfaye (2017) in Ethiopia

Objectives

- What are the **determinants** of infant mortality in India?

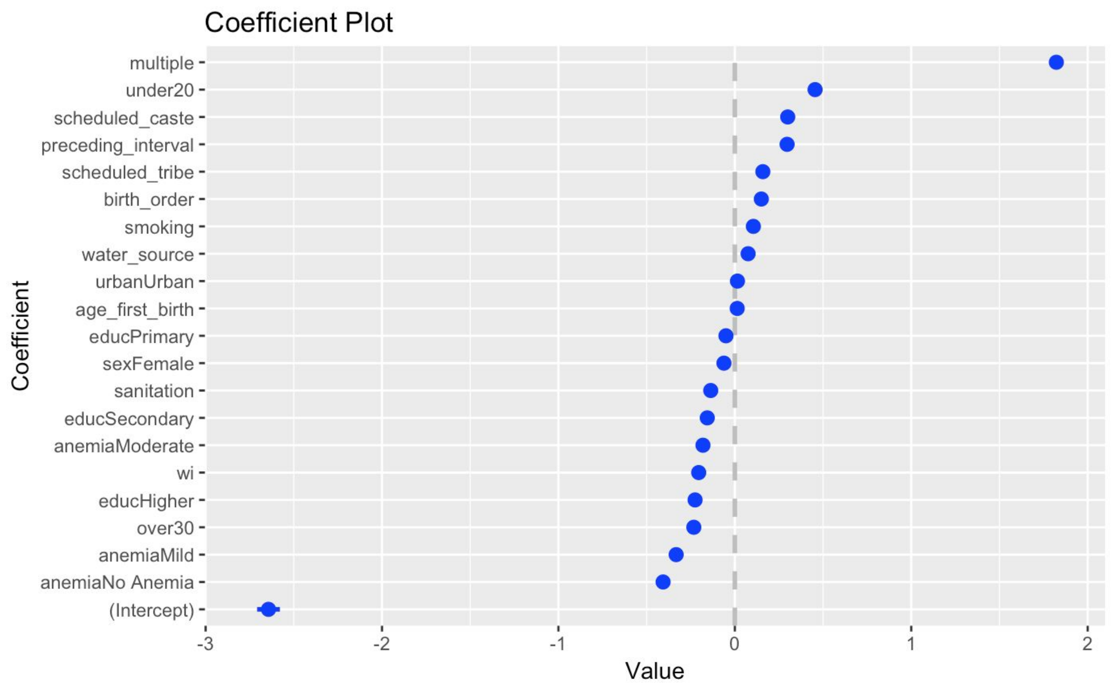
- Can we identify at-risk babies at the time of pregnancy **without medical expertise**?

Relevant Literature

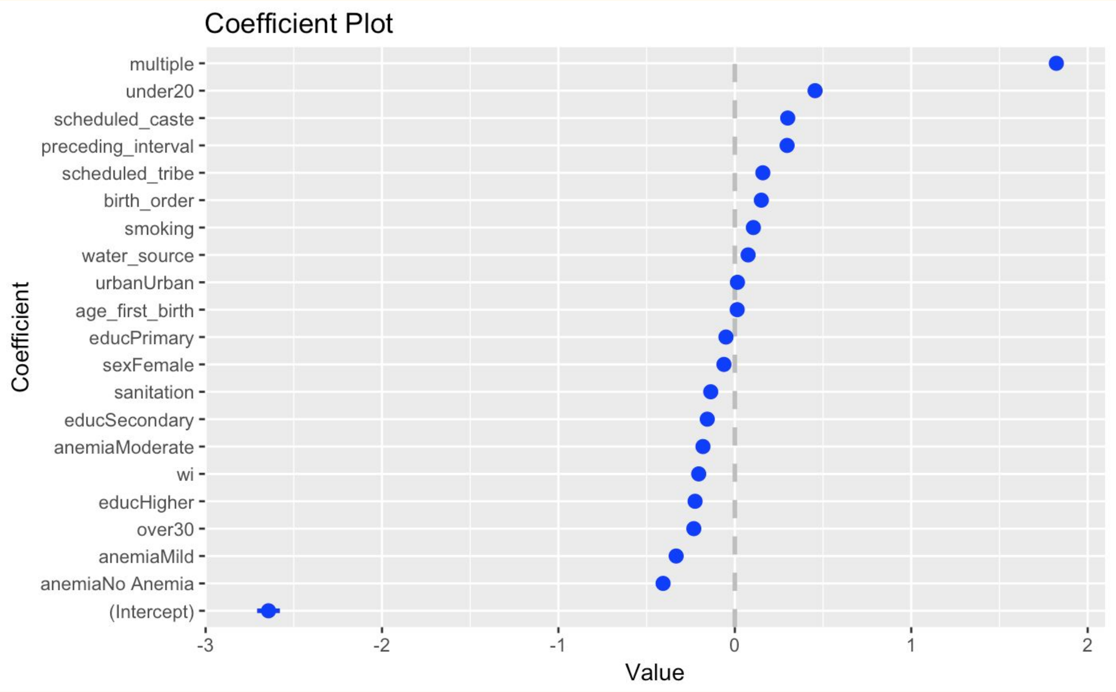
- Multiple studies have found tree-based methods to provide the best results for prediction of health outcomes (Song 2004, Carmelli 2007, Chen 2011, Tesfaye 2017)

- Most recent infant mortality prediction : Tesfaye (2017)
    - Accuracy : 90.38%

What are the determinants of infant mortality?



What are the determinants of infant mortality?



Prediction using Random Forest

- Accuracy - 92%
- Higher than Tesfaye (2017)
- Rejoice? No.
- We misclassified **3 out of 4** at-risk babies as being healthy!
- Focus on minimizing False Negatives
- Trade-off with False Positives
  - Resource constraints
- Class Imbalance Problem
  - 93:7

Prediction using Random Forest

- Accuracy - 92%
- Higher than Tesfaye (2017)
- Rejoice? No.
- We misclassified **3 out of 4** at-risk babies as being healthy!
- Focus on minimizing False Negatives
- Trade-off with False Positives
    - Resource constraints
- Class Imbalance Problem
    - 93:7

What can we do to address class imbalance?

- **Asymmetric penalties**
  - Higher penalties for misclassifications on the minority class
- **Oversample from the minority class**
  - Overfitting, emphasis on outliers
- **Random undersampling from the majority class**
  - Not robust to cross-validation, ‘wasting’ information

What can we do to address class imbalance?

- **Asymmetric penalties**
  - Higher penalties for misclassifications on the minority class

- **Oversample from the minority class**
  - Overfitting, emphasis on outliers

- **Random undersampling from the majority class**
  - Not robust to cross-validation, ‘wasting’ information

What can we do to address class imbalance?

- Asymmetric penalties
- Oversample from the minority class
- Random undersampling from the majority class
- **Synthetic Minority Oversampling (SMOTE)**
  - **Create synthetic observations from minority class**

- **Remove Tomek Links**

What are Tomek Links?

A pair of observations forms a Tomek Link if

- They are each other's **nearest neighbour** and
- They belong to **different classes**

- One of the observations is **noise** OR
- Both observations are **borderline**

We drop the observation belonging to the majority class.

What are Tomek Links?

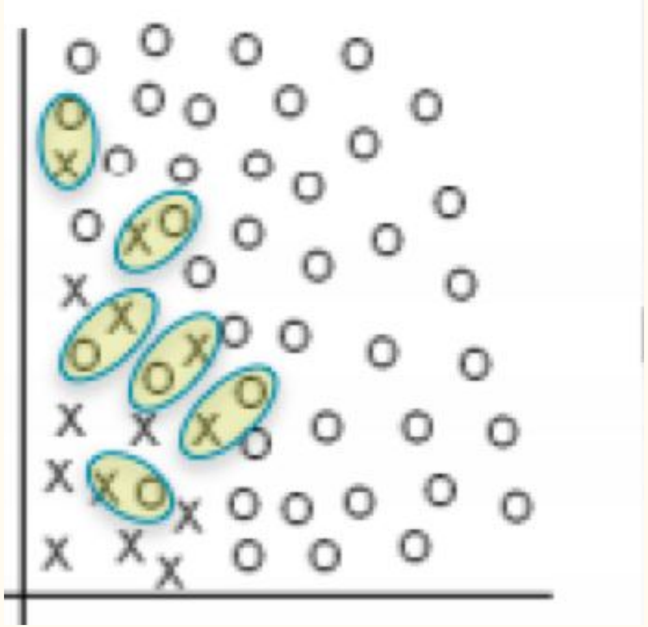
A pair of observations forms a Tomek Link if

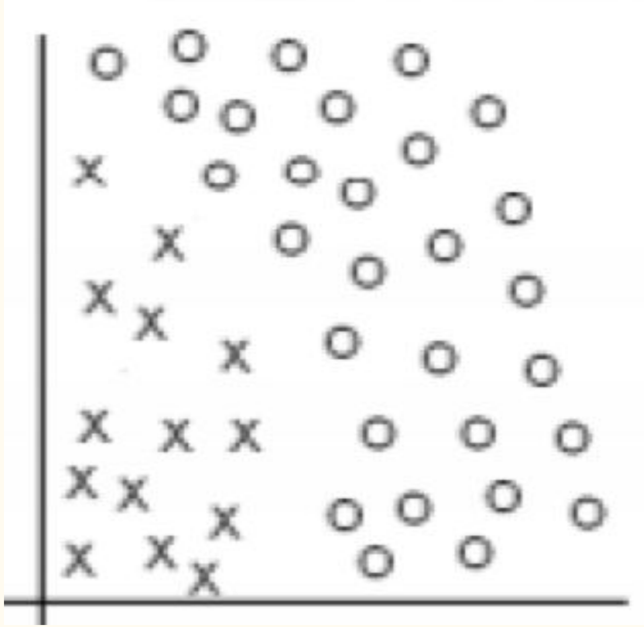
- They are each other’s **nearest neighbour** and
- They belong to **different classes**

- One of the observations is **noise** OR
- Both observations are **borderline**

We drop the observation belonging to the majority class.

Removing Tomek Links





Workflow

Clean Data

Recode variables, scale and standardize features to be centered around 0 with standard deviation 1

Resample

Apply a resampling method to address the class imbalance problem

Classify

Train an appropriate classifier to identify at-risk babies.

Tune hyper-parameters

Tune on the classifier’s hyperparameters to improve prediction.

Results

		Accuracy	False Negative Rate	False Positive Rate
Random Forest	*Original Data*	92%	75%	7%
	Without Tomek Links	92%	63%	8%
Logistic Regression	*Original Data*	93%	51%	7%
	Without Tomek Links	93%	**37%**	8%

Results

- Tomek links help distinguish between classes
- In each classifier, removing Tomek links improves accuracy on the minority class by at least 15%
- Logistic Regression does a better job than Random Forests

Further Work

- Cross-validation
- Combinations of resampling methods
  - Undersampling + Tomek
- Other classifiers
  - AdaBoost
  - GradientBoosting
- Explore why Logistic Regression predicts so much better than Random Forest

Further Work

- Cross-validation

- Combinations of resampling methods
  - Undersampling + Tomek

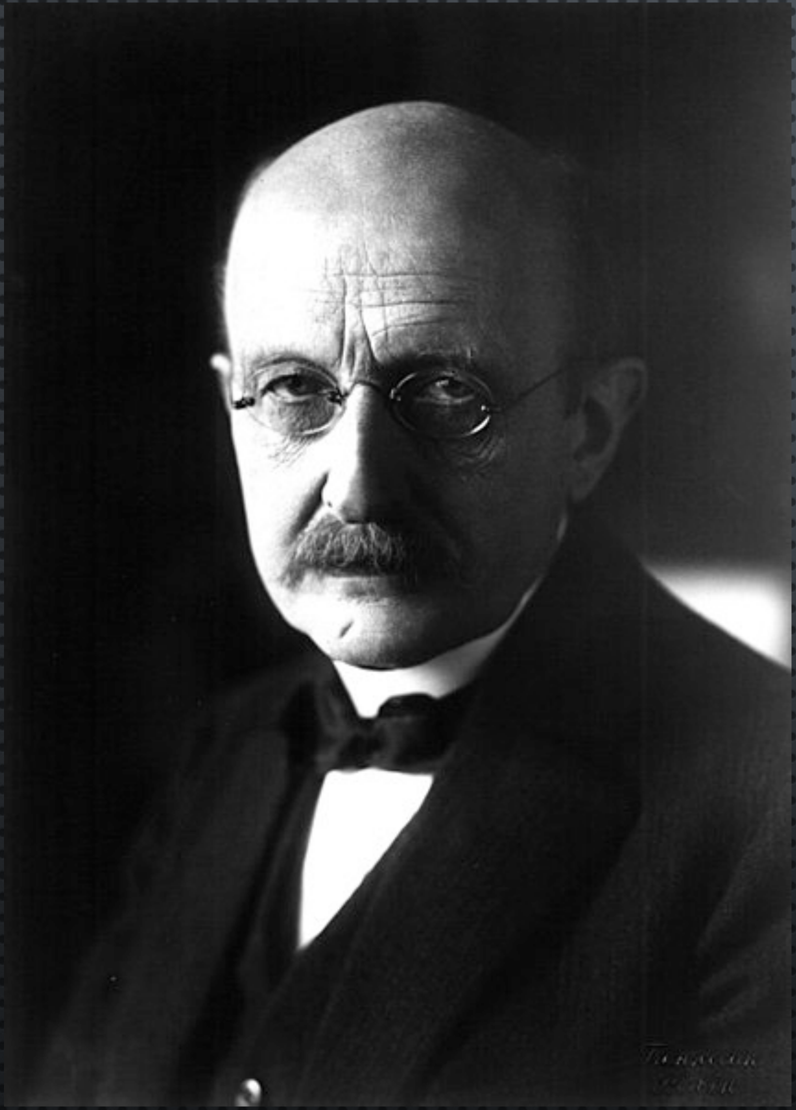
- Other classifiers
  - AdaBoost
  - GradientBoosting

- Explore why Logistic Regression predicts so much better than Random Forest

Speaker Transition

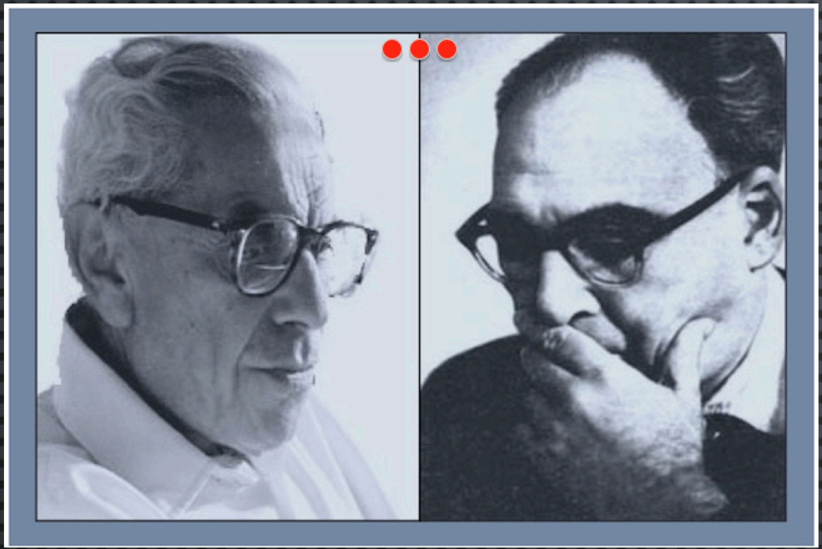
STANDING ON THE SHOULDERS OF GIANTS: UNDERSTANDING THE EVOLUTION OF SCIENCE

RODRIGO VALDÉS ORTIZ



Max Planck

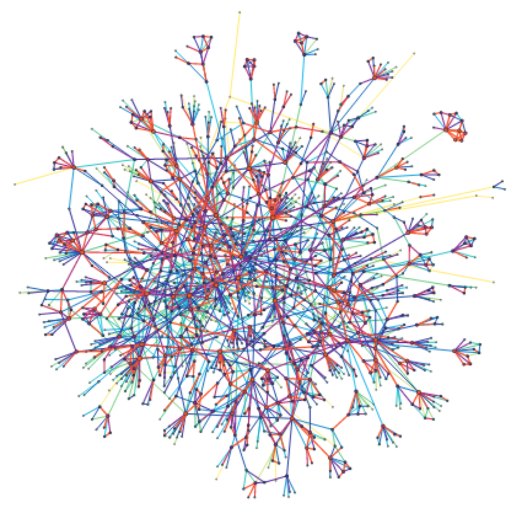
Peter Higgs

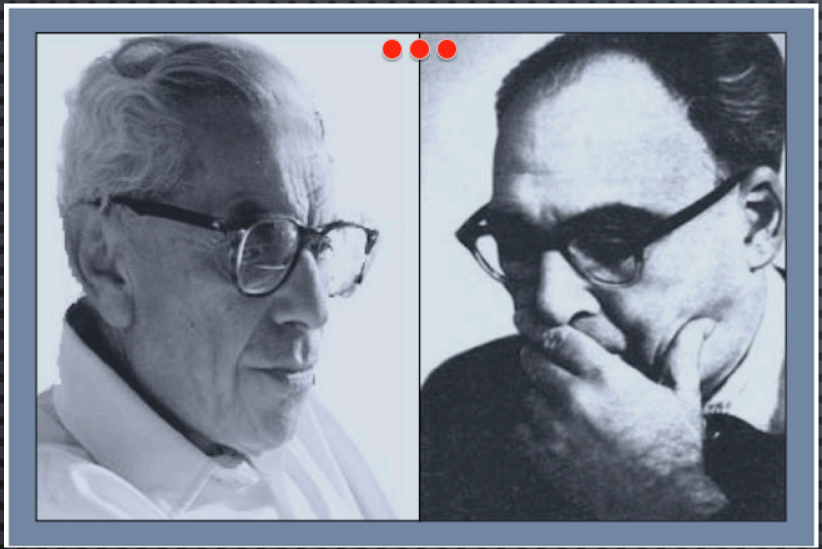


Erdos and Renyi

Barabási

Granovetter

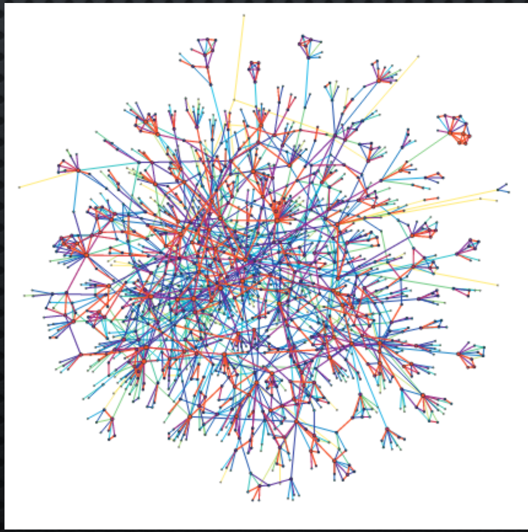


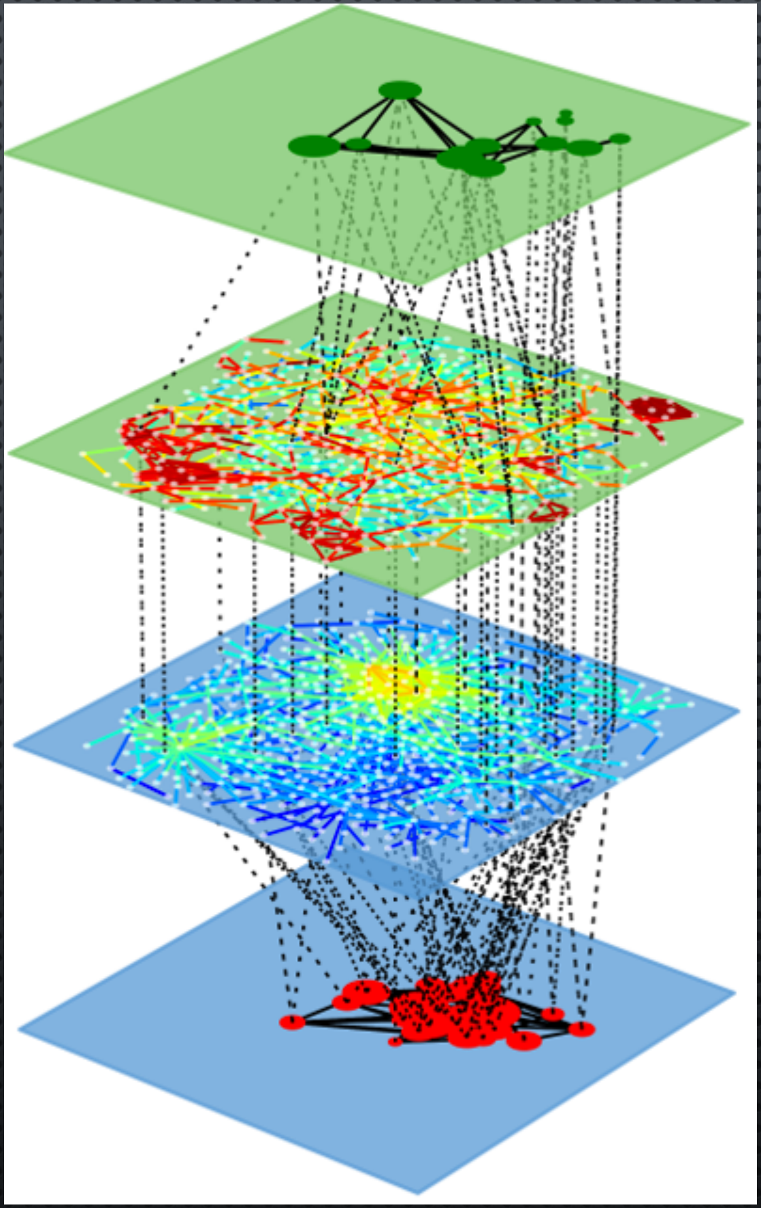


Erdos and Renyi

Barabási

Granovetter





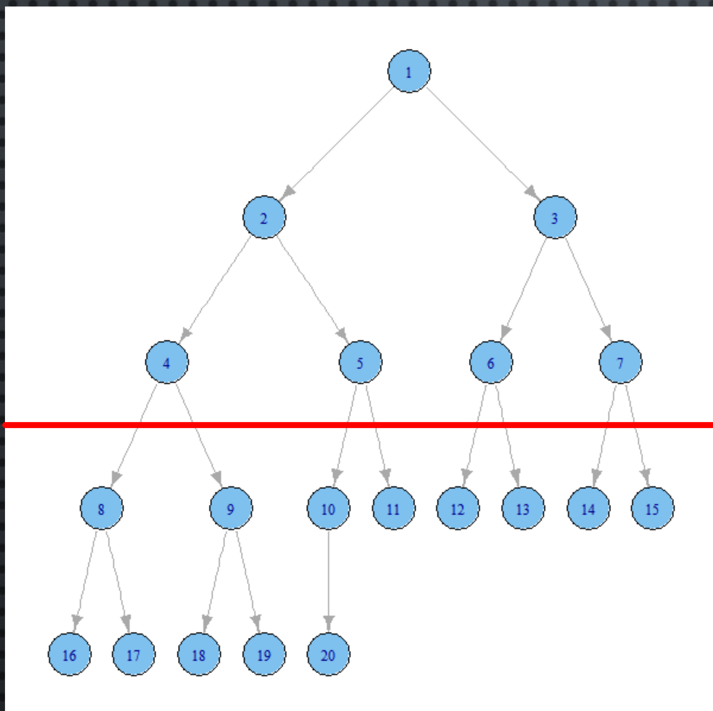
Sorce: MuxViz samples.

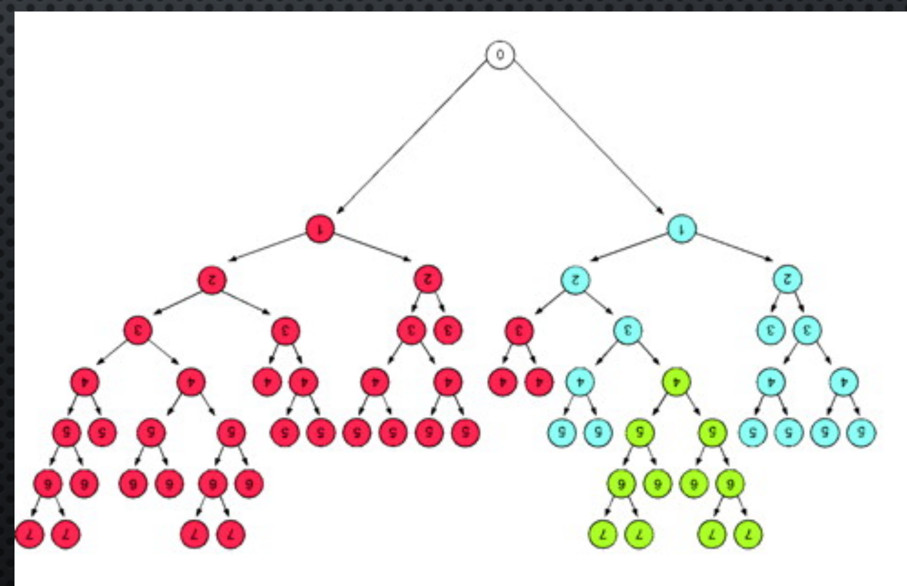
Build networks with Infomap

Groups

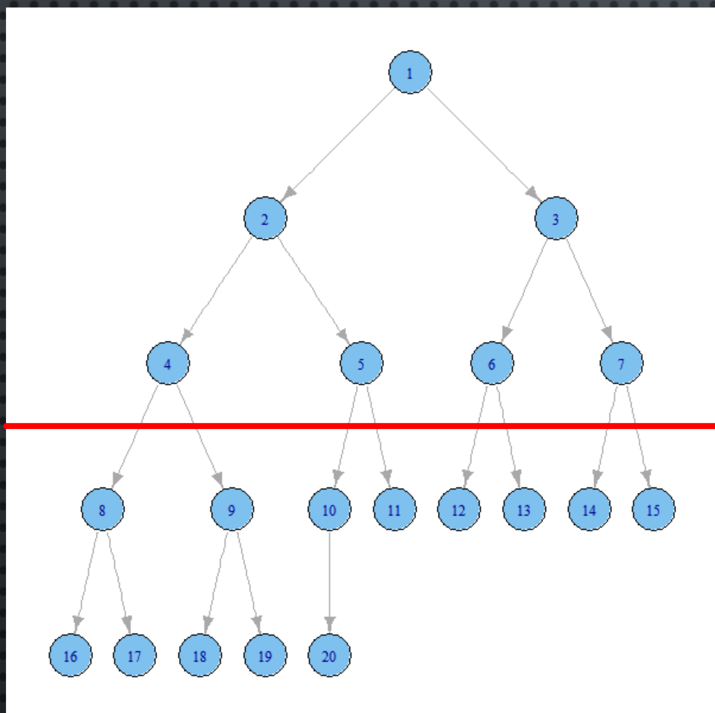
Relationships

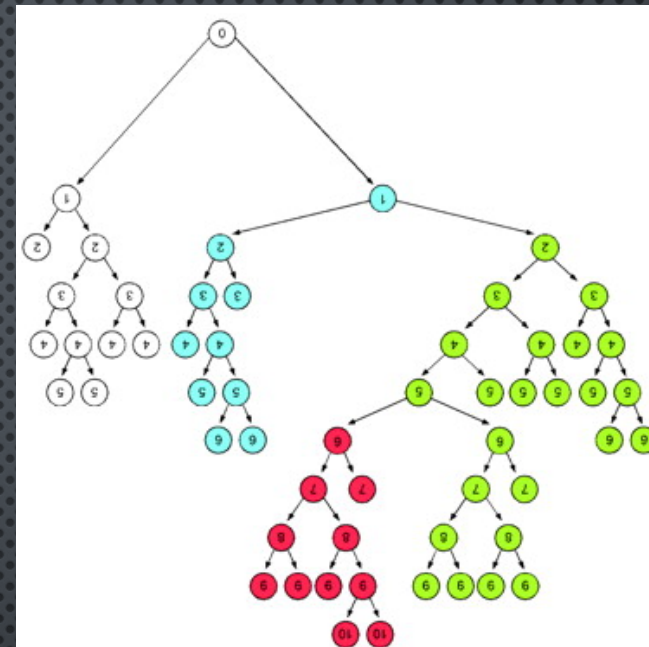
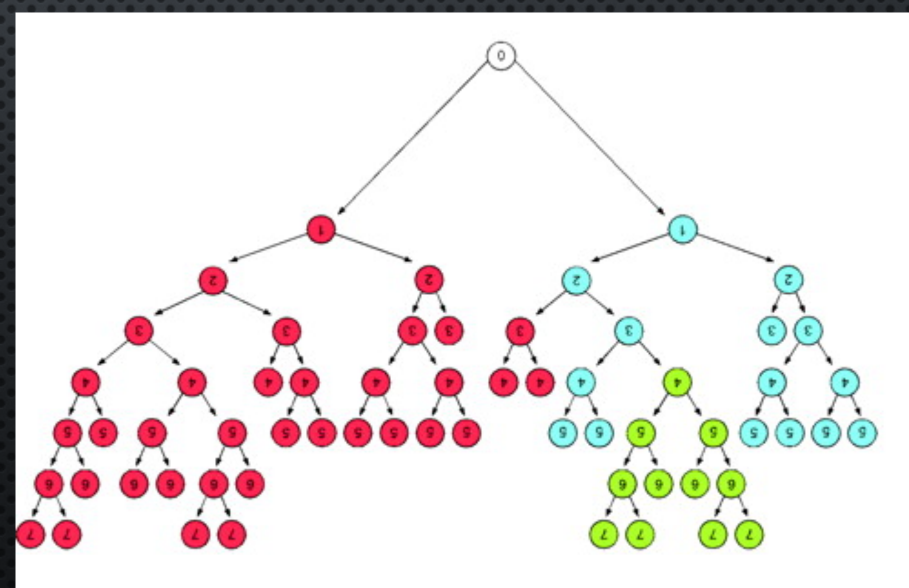
Infomap software: Edler and Rosvall





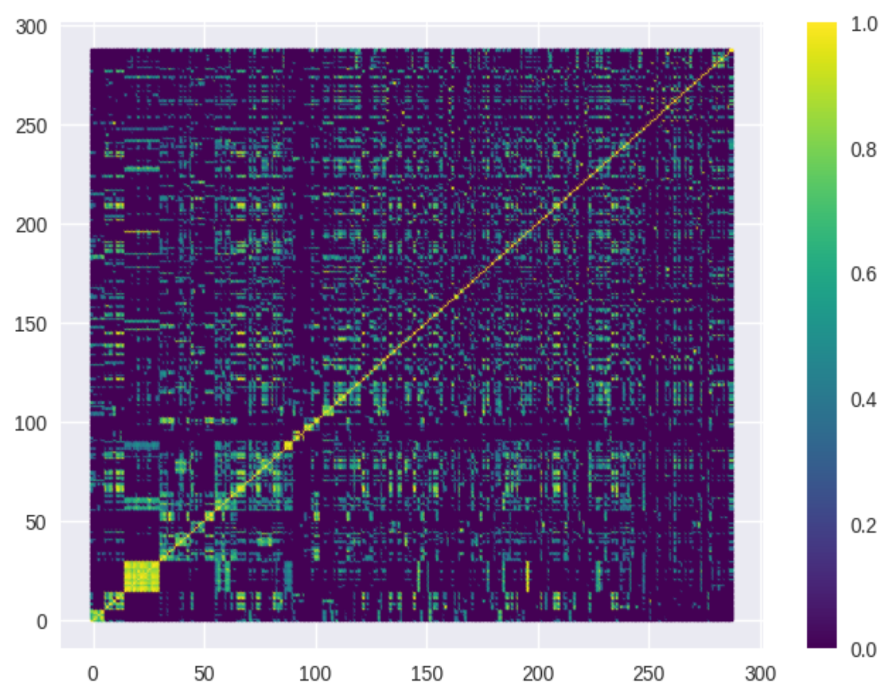
Sorce: Stackoverflow.



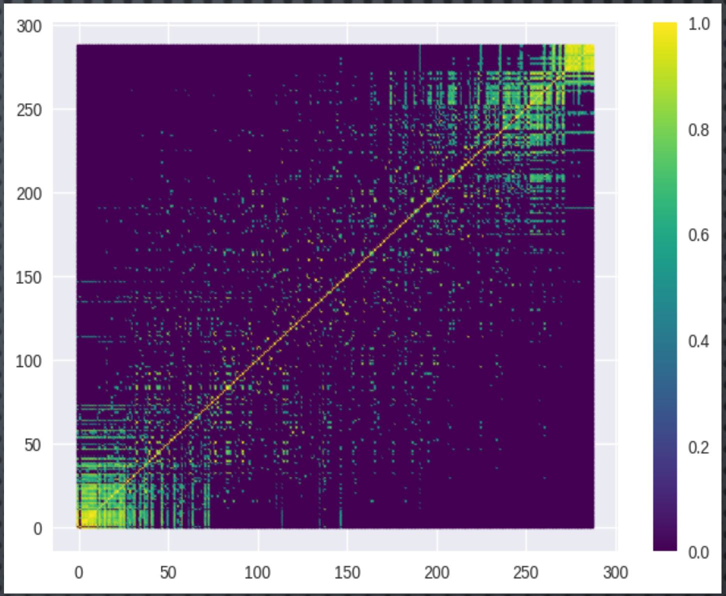



Sorce: Stackoverflow.

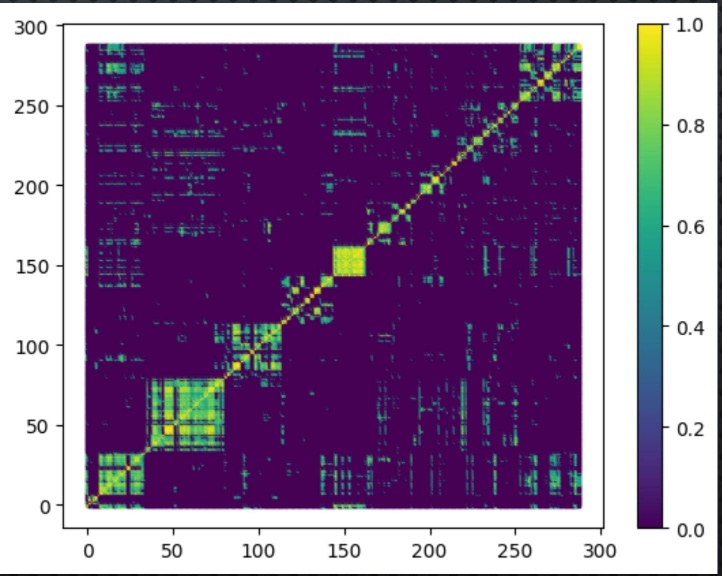
Correlation among groups:
Medicine/Bio Cluster



Baseline (output Infomap)



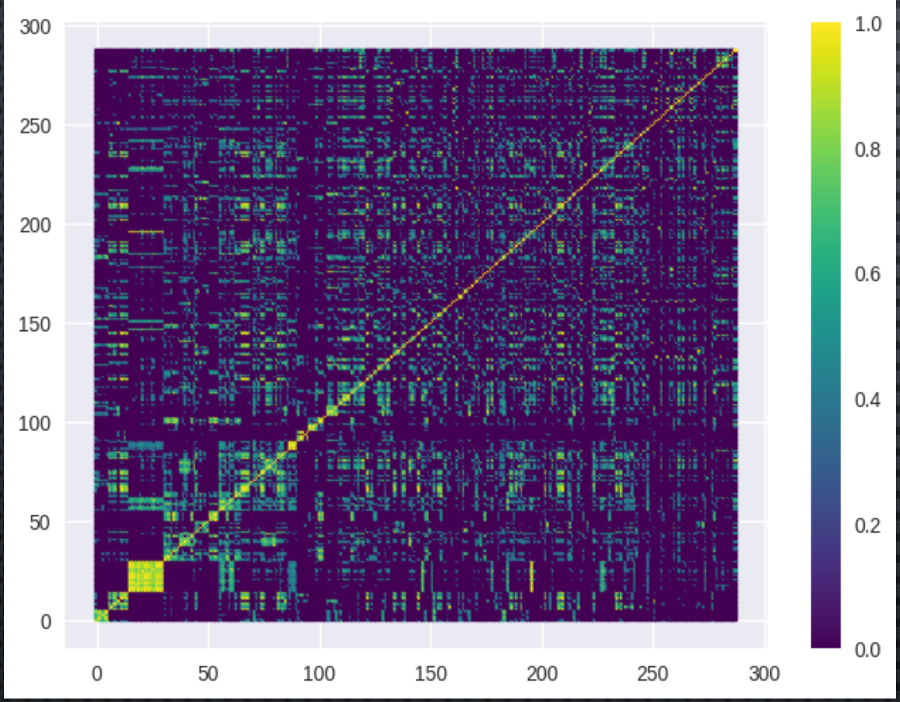
PCA



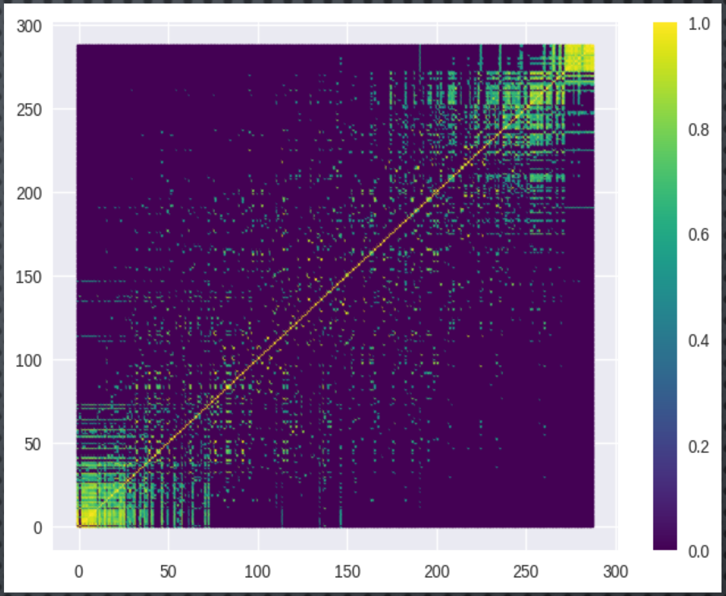
Recursive TSNE

~ 830 total groups
~ 280 medicine/bio

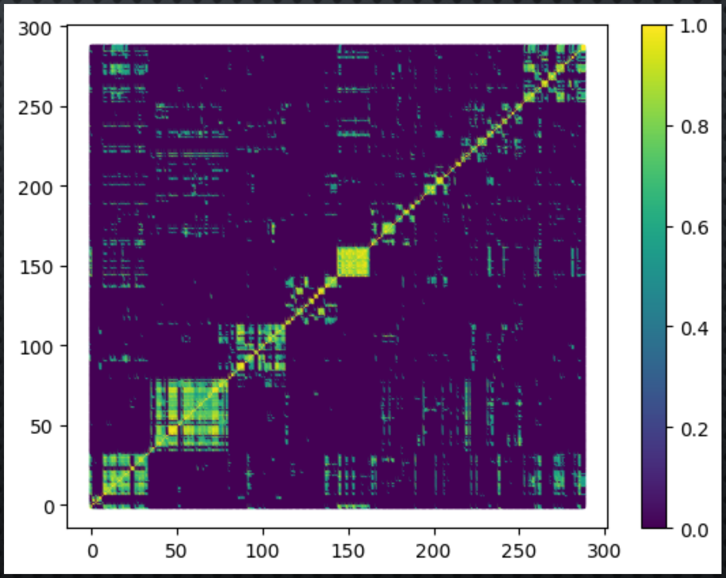
Correlation among groups:
Medicine/Bio Cluster



Baseline (output Infomap)



PCA

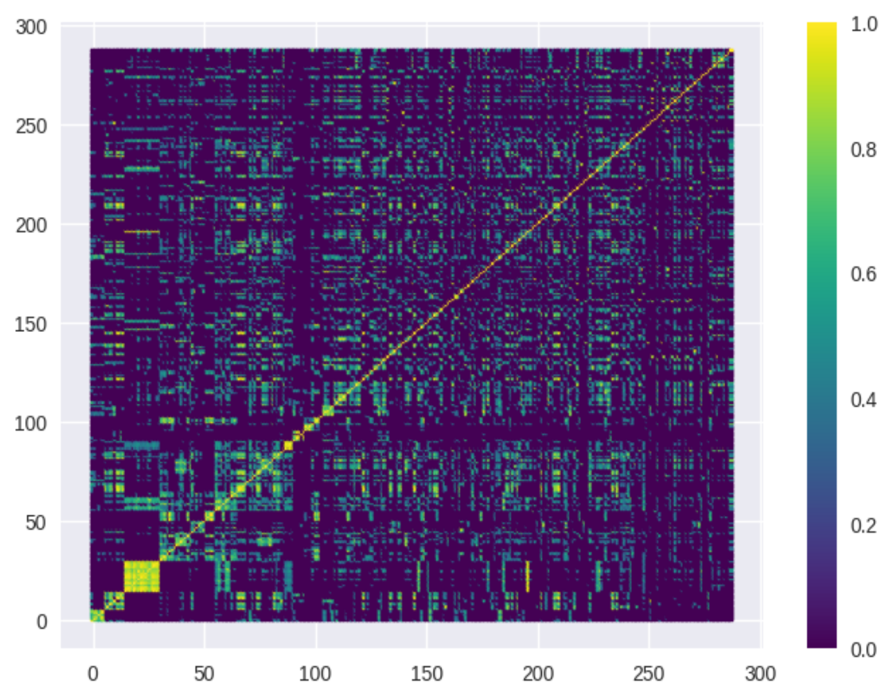


Recursive TSNE

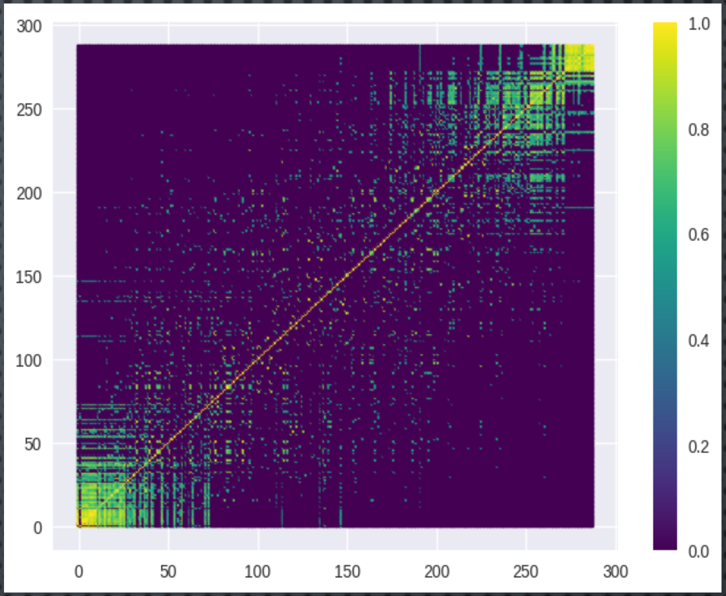
~ 830 total groups
~ 280 medicine/bio

Order: recursive algorithms
Assess: not only with eyes

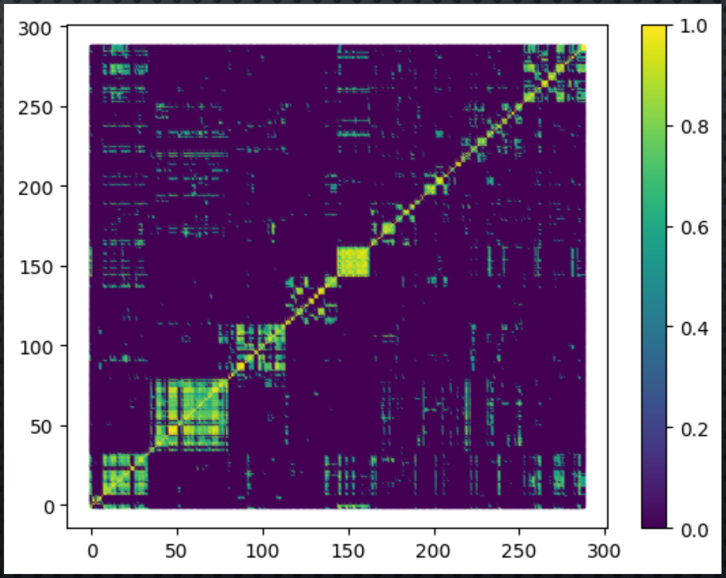
Correlation among groups:
Medicine/Bio Cluster



Baseline (output Infomap)



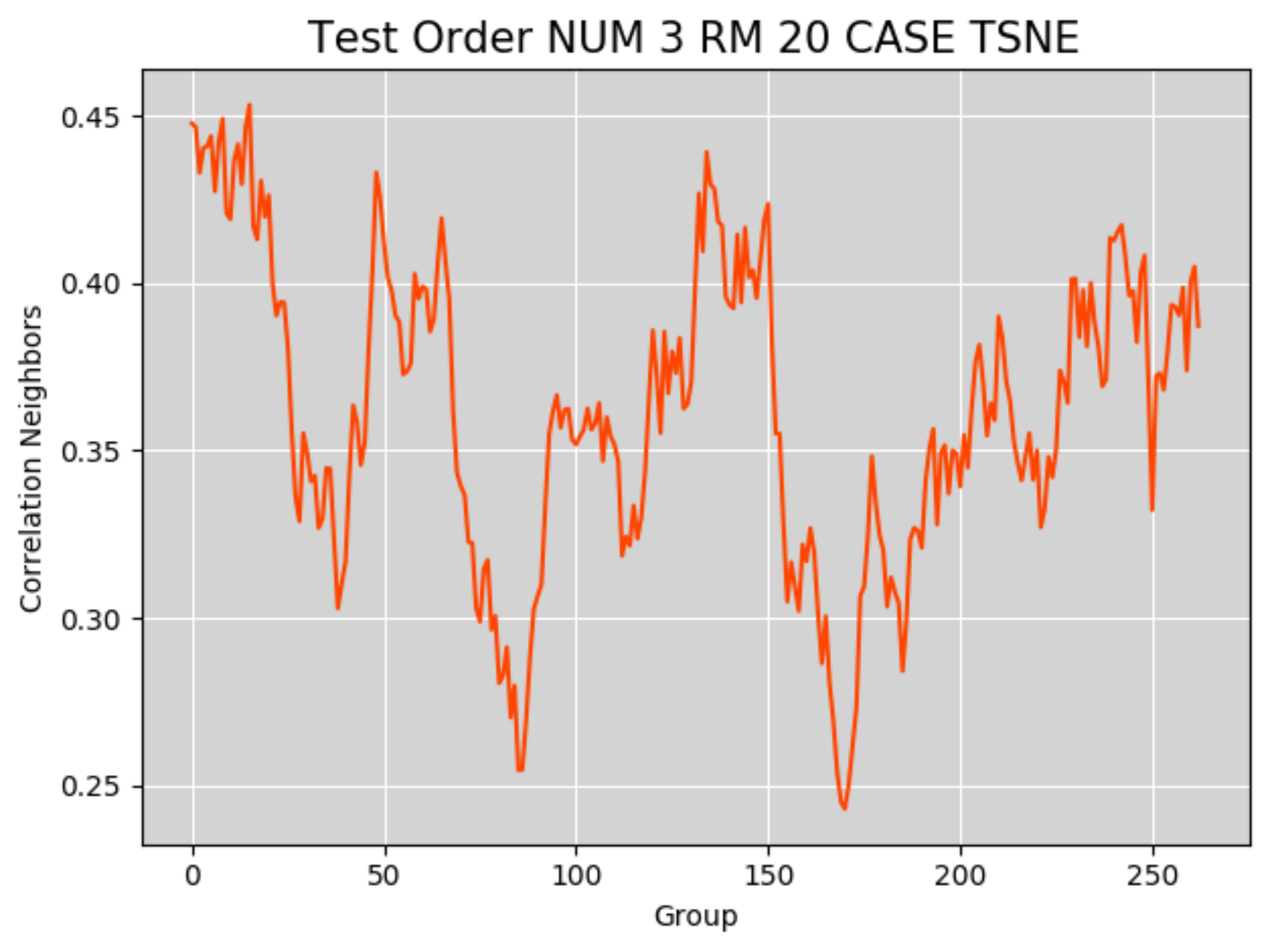
PCA



Recursive TSNE

~ 830 total groups
~ 280 medicine/bio

Order: recursive algorithms
Assess: not only with eyes


Test Order NUM 3 RM 20 CASE TSNE
Correlation Neighbors
0.45
0.40
0.35
0.30
0.25
0
50
100
150
200
250
Group

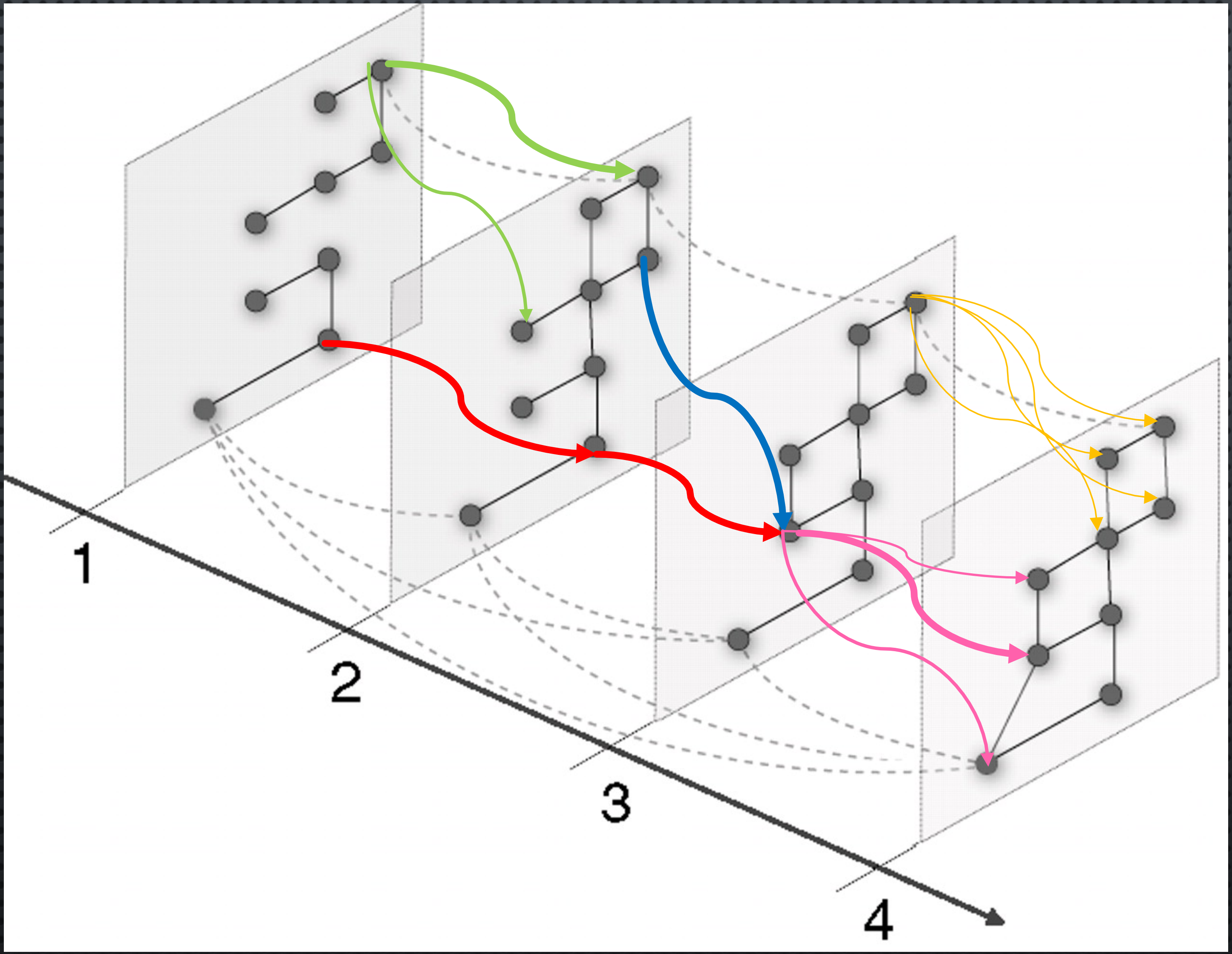


Image from Peter J. Mucha et al., Science, 2010;328:876-878

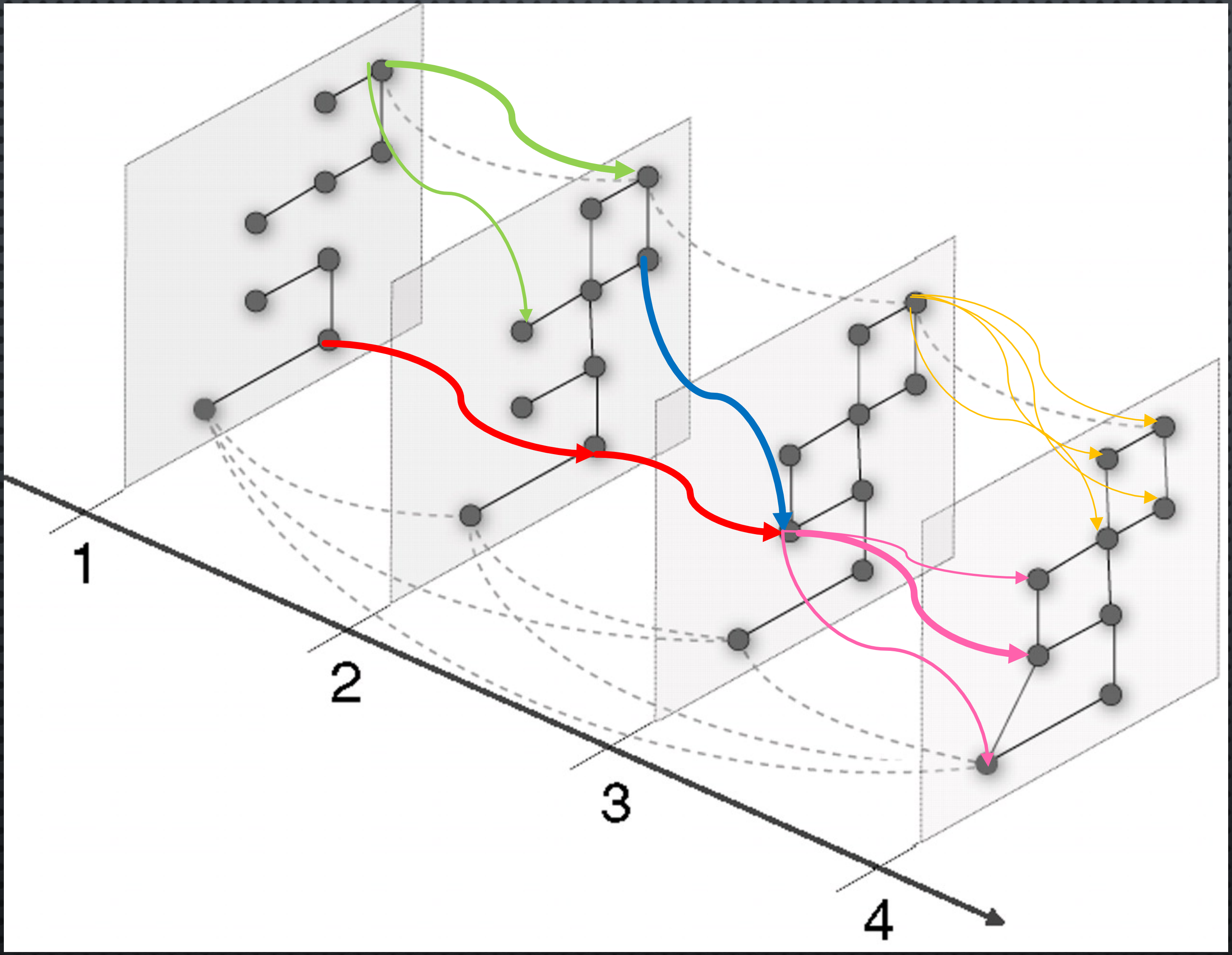


Image from Peter J. Mucha et al., Science, 2010;328:876-878

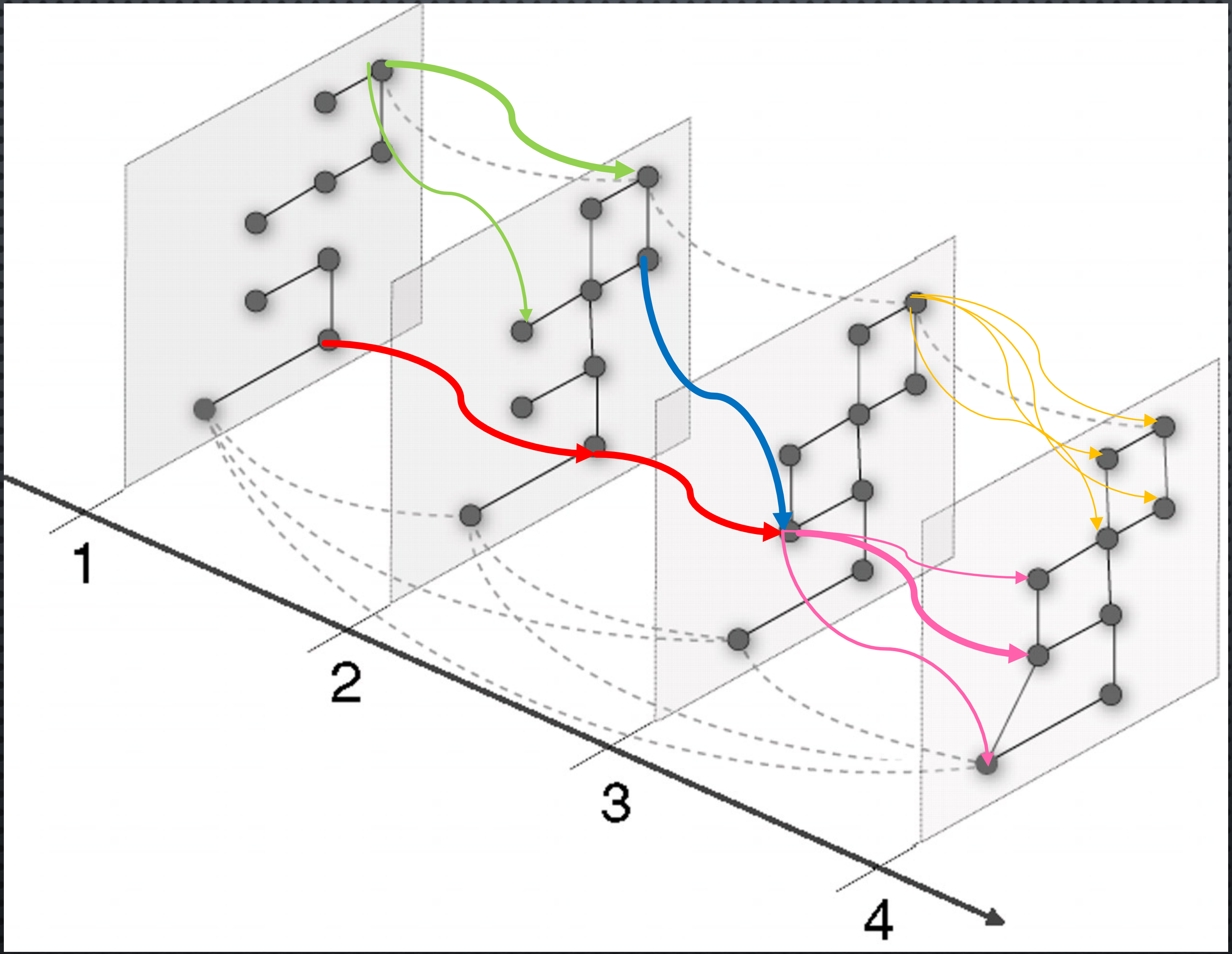


Image from Peter J. Mucha et al., Science, 2010;328:876-878

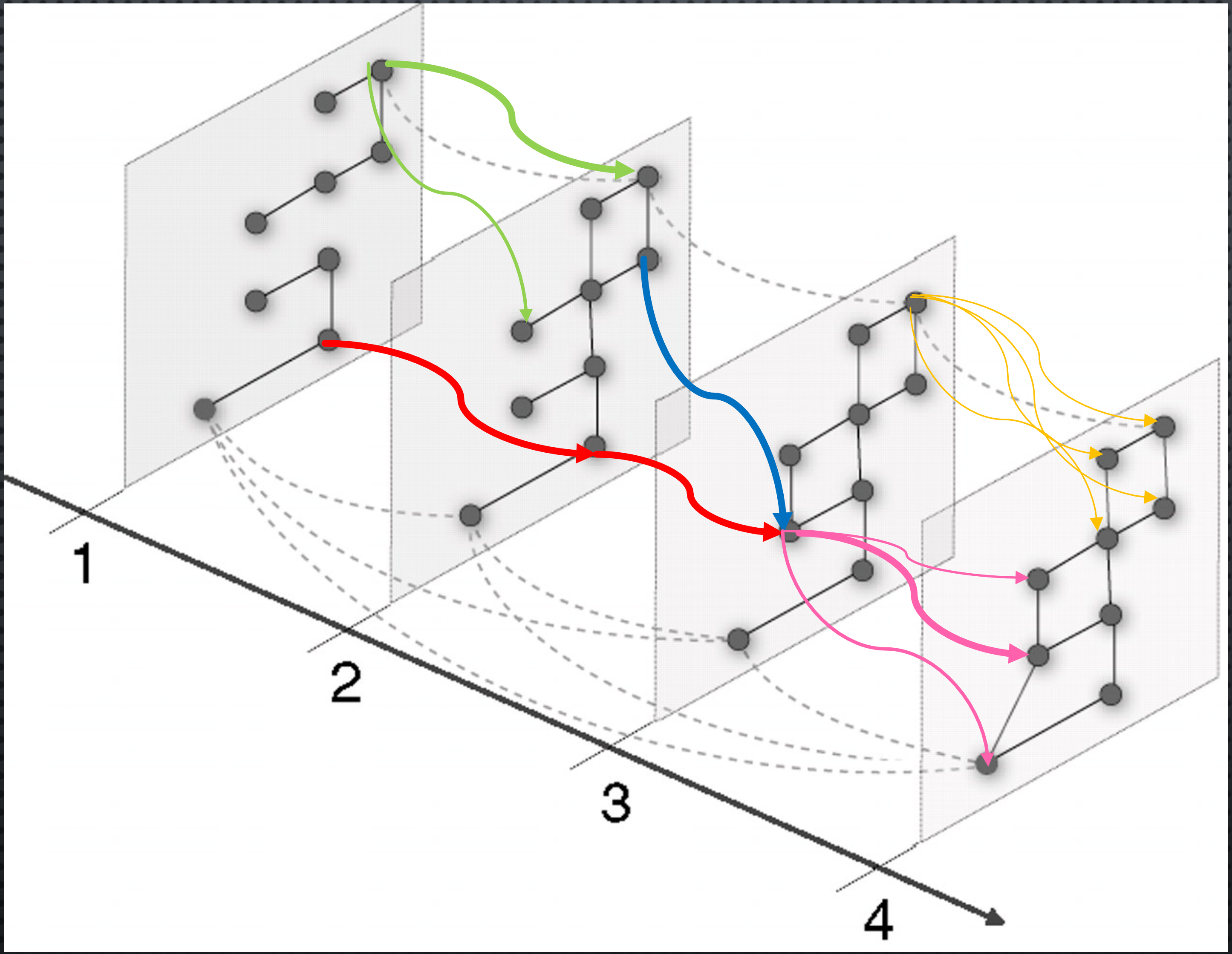
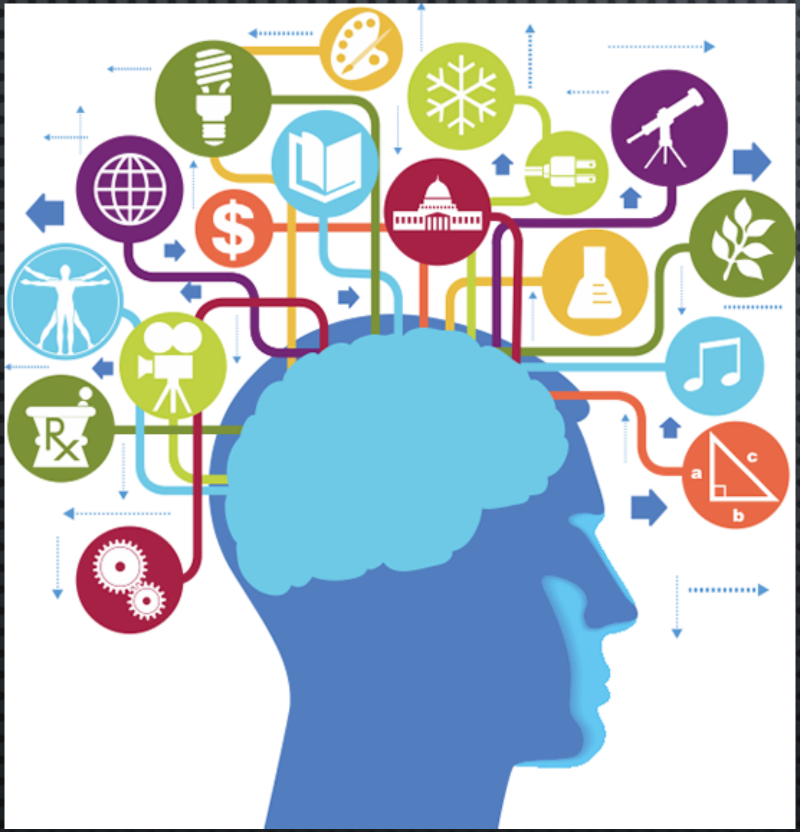
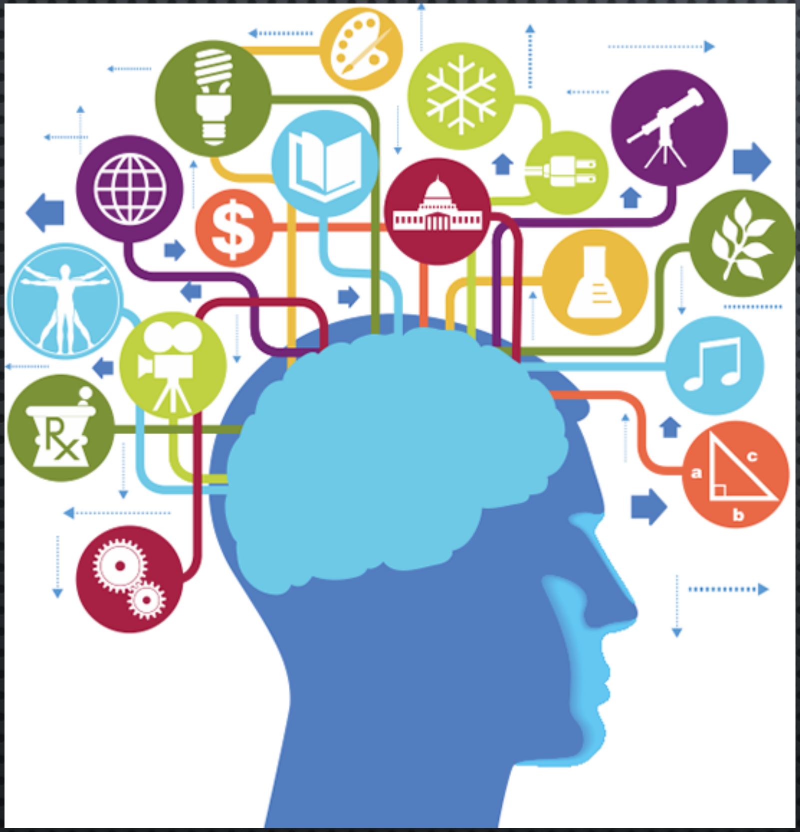
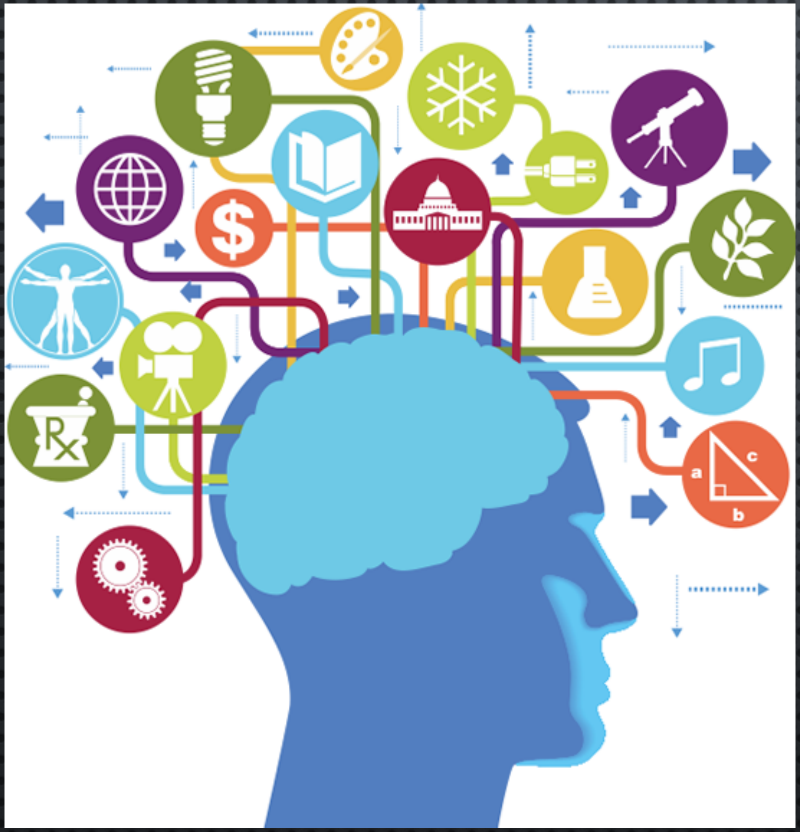


Image from Peter J. Mucha et al., Science, 2010;328:876-878





Speaker Transition

Estimation of Income Process and Heterogeneity in Partial Insurance: A Quantile Regression Based Auxiliary Particle Filter Approach

Luxi Han
Computational Social Science
University of Chicago

2018.03.28

Background: Income Process

$$Y_{i,t} = \xi_{i,t} + \epsilon_{i,t}$$
$$\xi_{i,t} = \xi_{i,t} + \eta_{i,t}$$

$$\text{s.t. Permanent:} \quad \xi_{i,t}$$
$$\text{Transitory:} \quad \epsilon_{i,t}$$
$$(\eta_{i,t}, \epsilon_{i,t})\tilde{}F(\eta_{i,t}, \epsilon_{i,t})$$

Background: Income Process

$$Y_{i,t} = \xi_{i,t} + \epsilon_{i,t}$$
$$\xi_{i,t} = \xi_{i,t} + \eta_{i,t}$$

$$\text{s.t. Permanent:} \quad \xi_{i,t}$$
$$\text{Transitory:} \quad \epsilon_{i,t}$$
$$(\eta_{i,t}, \epsilon_{i,t})\tilde{}F(\eta_{i,t}, \epsilon_{i,t})$$

Background: Income Process

$$Y_{i,t} = \xi_{i,t} + \epsilon_{i,t}$$
$$\xi_{i,t} = \xi_{i,t} + \eta_{i,t}$$

$$\text{s.t. Permanent:} \quad \xi_{i,t}$$
$$\text{Transitory:} \quad \epsilon_{i,t}$$
$$(\eta_{i,t}, \epsilon_{i,t}) \tilde{} F(\eta_{i,t}, \epsilon_{i,t})$$

Research Question

1. What is the joint distribution of permanent and transitory income shocks
2. What is the overall partial insurance coefficient?
3. What is the distribution of consumption given η, ϵ, a?
4. How does the insurability vary over η, ϵ, a?
5. Does complete and incomplete information (DGP and shock components) change the consumption rule?

Research Question

1. What is the joint distribution of permanent and transitory income shocks
2. What is the overall partial insurance coefficient?
3. What is the distribution of consumption given η, ϵ, a?
4. How does the insurability vary over η, ϵ, a?
5. Does complete and incomplete information (DGP and shock components) change the consumption rule?

Data

1. Major Data Source: SIAB Scientific Use File
2. Alternative: PSID and CEX

Data

1. Major Data Source: SIAB Scientific Use File
2. Alternative: PSID and CEX

Estimation of Income Process: Auxiliary Particle Filter Approach

EM algorithm follows the following steps:

1. E Step: Draw particles from the posterior: $f(\xi_{i,t}|y_{i,1:T})$ using two filter smoother
2. M Step: Estimate parameters of interest using quantile regressions:

$$Q_j(x) = \begin{cases} x_0\beta_0; & j = j^* \\ x_0\beta_0 + \sum_{k=j+1}^{J} exp(x'\beta_k); & j > j^* \\ x_0\beta_0 - \sum_{k=1}^{j-1} exp(x'\beta_k); & j < j^* \end{cases}$$

Estimation of Income Process: Auxiliary Particle Filter Approach

EM algorithm follows the following steps:

1. E Step: Draw particles from the posterior: $f(\xi_{i,t}|y_{i,1:T})$ using two filter smoother
2. M Step: Estimate parameters of interest using quantile regressions:

$$Q_j(x) = \begin{cases} x_0\beta_0; & j = j^* \\ x_0\beta_0 + \sum_{k=j+1}^{J} exp(x'\beta_k); & j > j^* \\ x_0\beta_0 - \sum_{k=1}^{j-1} exp(x'\beta_k); & j < j^* \end{cases}$$

Estimation of Income Process: Auxiliary Particle Filter Approach

EM algorithm follows the following steps:

1. E Step: Draw particles from the posterior: $f(\xi_{i,t}|y_{i,1:T})$ using two filter smoother
2. M Step: Estimate parameters of interest using quantile regressions:

$$Q_j(x) = \begin{cases} x_0\beta_0; & j = j^* \\ x_0\beta_0 + \sum_{k=j+1}^{J} exp(x'\beta_k); & j > j^* \\ x_0\beta_0 - \sum_{k=1}^{j-1} exp(x'\beta_k); & j < j^* \end{cases}$$

Estimation of Income Process: Auxiliary Particle Filter Approach

EM algorithm follows the following steps:

1. E Step: Draw particles from the posterior: $f(\xi_{i,t}|y_{i,1:T})$ using two filter smoother
2. M Step: Estimate parameters of interest using quantile regressions:

$$Q_j(x) = \begin{cases} x_0\beta_0; \quad j = j^* \\ x_0\beta_0 + \sum_{k=j+1}^{J} exp(x'\beta_k); \quad j > j^* \\ x_0\beta_0 - \sum_{k=1}^{j-1} exp(x'\beta_k); \quad j < j^* \end{cases}$$

Estimation of Partial insurance

Key Variables:

$$f(C_{i,t}|\xi_{i,t},\epsilon_{i,t},a_{i,t})$$
$$f(\frac{\partial E[C_{i,t}|\xi_{i,t},\epsilon_{i,t},a_{i,t}]}{\partial \eta_t})$$
$$f(\frac{\partial E[C_{i,t}|\xi_{i,t},\epsilon_{i,t},a_{i,t}]}{\partial \epsilon_t})$$

Estimation of Partial insurance

Key Variables:

$$f(C_{i,t}|\xi_{i,t}, \epsilon_{i,t}, a_{i,t})$$
$$f(\frac{\partial E[C_{i,t}|\xi_{i,t}, \epsilon_{i,t}, a_{i,t}]}{\partial \eta_t})$$
$$f(\frac{\partial E[C_{i,t}|\xi_{i,t}, \epsilon_{i,t}, a_{i,t}]}{\partial \epsilon_t})$$

Estimation of Partial insurance

1. Route 1: Consumption Rule from a PIC model (Carroll, 1997) given estimated income process

$$\max_{c_t, a_t} \sum_{t=1}^{T} u(c_t)$$
$$s.t. c_t + a_{t+1} = y_t + (1+r)a_t$$
$$y_t = \xi_t + \epsilon_t$$
$$(\eta_{i,t}, \epsilon_{i,t}) \tilde{} F(\eta_{i,t}, \epsilon_{i,t})$$
$$a_t > 0 \forall t$$

2. Estimation procedure follows from Monte Carlo, or rerun the EM algorithm on consumption (Arellano, Blundell, Bonohomme, 2016)

Estimation of Partial insurance

1. Route 1: Consumption Rule from a PIC model (Carroll, 1997) given estimated income process

$$\max_{c_t, a_t} \sum_{t=1}^{T} u(c_t)$$
$$s.t. c_t + a_{t+1} = y_t + (1+r)a_t$$
$$y_t = \xi_t + \epsilon_t$$
$$(\eta_{i,t}, \epsilon_{i,t})\tilde{} F(\eta_{i,t}, \epsilon_{i,t})$$
$$a_t > 0 \forall t$$

2. Estimation procedure follows from Monte Carlo, or rerun the EM algorithm on consumption (Arellano, Blundell, Bonohomme, 2016)

Estimation of Partial insurance

1. Route 1: Consumption Rule from a PIC model (Carroll, 1997) given estimated income process

$$\max_{c_t, a_t} \sum_{t=1}^{T} u(c_t)$$

$$s.t. c_t + a_{t+1} = y_t + (1+r)a_t$$

$$y_t = \xi_t + \epsilon_t$$

$$(\eta_{i,t}, \epsilon_{i,t})\tilde{} F(\eta_{i,t}, \epsilon_{i,t})$$

$$a_t > 0 \forall t$$

2. Estimation procedure follows from Monte Carlo, or rerun the EM algorithm on consumption (Arellano, Blundell, Bonohomme, 2016)

Estimation of Partial insurance

1. Route 2: rerun the EM algorithm on consumption data in PSID and CEX

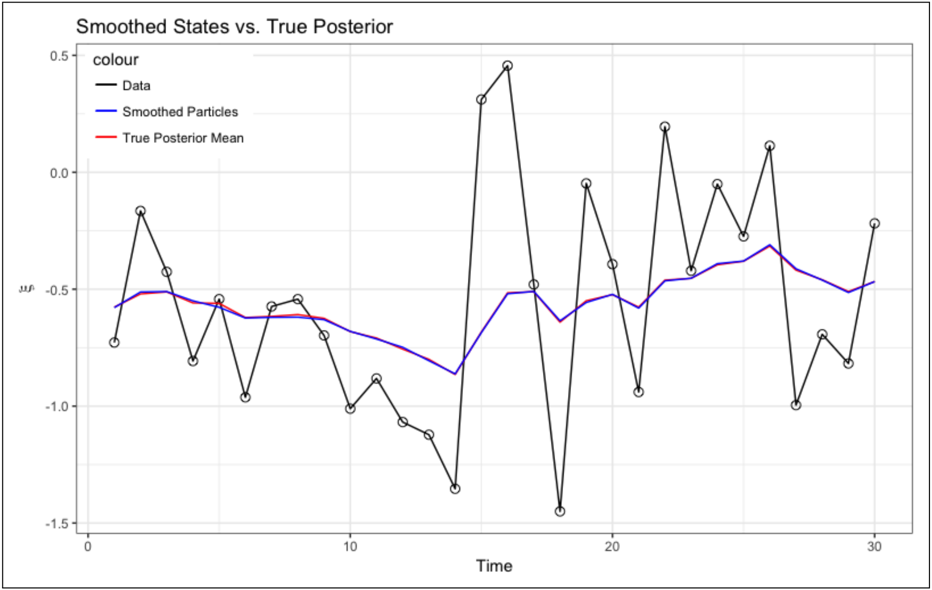


Figure: Smoothed States vs. True Posterior

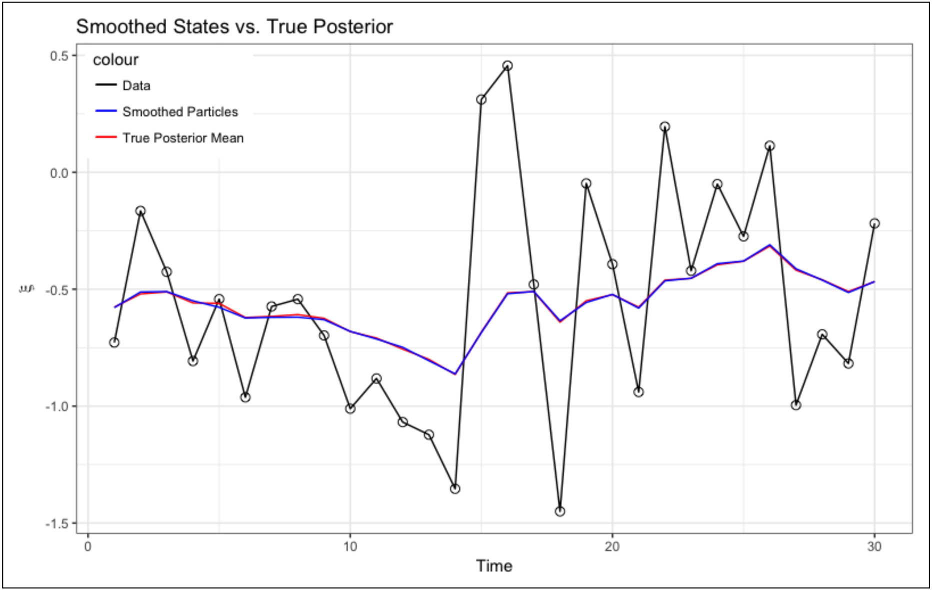


Figure: Smoothed States vs. True Posterior

Speaker Transition

A Study about Sub-content Generation Structure

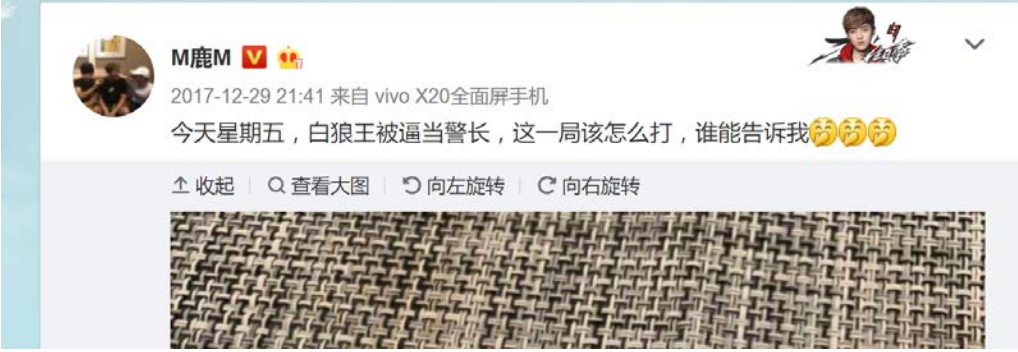
Yang Hou

A Study about Sub-content Generation Structure

Yang Hou

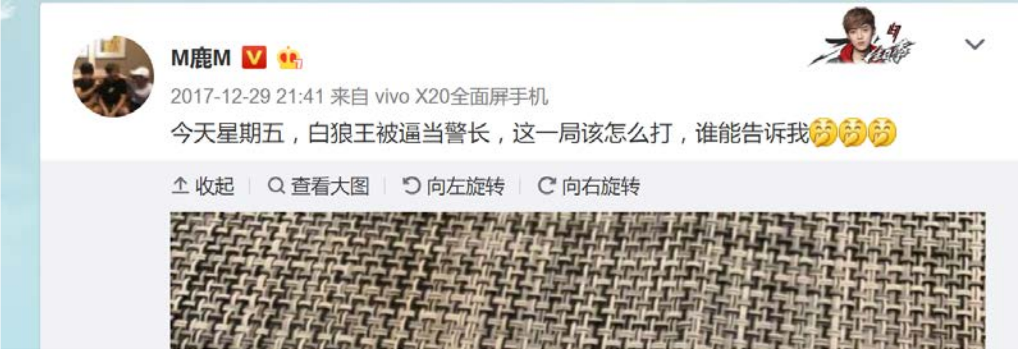
Background

A lot of sub-content is created under the post, combining followers' own experiences or emotions, provides a larger possibility to attract other followers.



Background

A lot of sub-content is created under the post, combining followers' own experiences or emotions, provides a larger possibility to attract other followers.



Subject

- Subject A: Han Lu
  - 1990-2012
  - 43.4 million followers
  - Average # of comments: 150 thousand
- Subject B: Yifan Wu
  - 1990-2012
  - 29.04 million followers
  - Average # of comments: 100 thousand
- Subject C: Geng Han
  - 1984-2005
  - 51.04 million followers
  - Average # of comments: 10 thousand

Post

- Only about personal life
- Every two months from 2017

Subject

- Subject A: Han Lu
  - 1990-2012
  - 43.4 million followers
  - Average # of comments: 150 thousand
- Subject B: Yifan Wu
  - 1990-2012
  - 29.04 million followers
  - Average # of comments: 100 thousand
- Subject C: Geng Han
  - 1984-2005
  - 51.04 million followers
  - Average # of comments: 10 thousand

Post

- Only about personal life
- Every two months from 2017

Subject

- Subject A: Han Lu
  - 1990-2012
  - 43.4 million followers
  - Average # of comments: 150 thousand
- Subject B: Yifan Wu
  - 1990-2012
  - 29.04 million followers
  - Average # of comments: 100 thousand
- Subject C: Geng Han
  - 1984-2005
  - 51.04 million followers
  - Average # of comments: 10 thousand

Post

- Only about personal life
- Every two months from 2017

Subject

- Subject A: Han Lu
  - 1990-2012
  - 43.4 million followers
  - Average # of comments: 150 thousand
- Subject B: Yifan Wu
  - 1990-2012
  - 29.04 million followers
  - Average # of comments: 100 thousand
- Subject C: Geng Han
  - 1984-2005
  - 51.04 million followers
  - Average # of comments: 10 thousand

Post

- Only about personal life
- Every two months from 2017

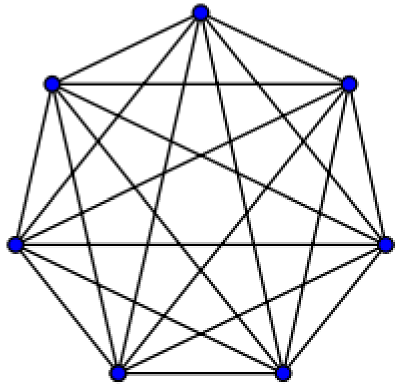
Network Forming Strategy

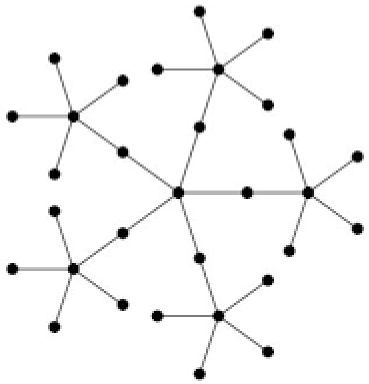
- Nodes: individual follower
- Edge Formation: when one individual replies to another's comments under the post
- Since every first level comments are all replying to the celebrity, the celebrity's node is omitted in the analysis to focus on the sub-content generation and sharing process among the followers and tourists.
- Comments with at least one like

Network Forming Strategy

- Nodes: individual follower
- Edge Formation: when one individual replies to another's comments under the post
- Since every first level comments are all replying to the celebrity, the celebrity's node is omitted in the analysis to focus on the sub-content generation and sharing process among the followers and tourists.
- Comments with at least one like

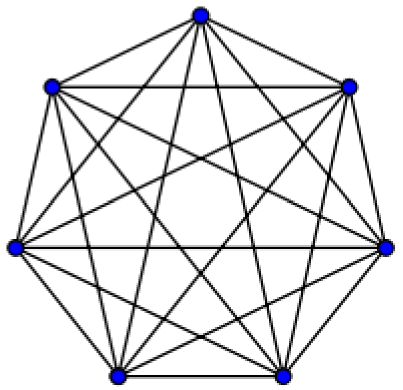
Extreme Case Discussion

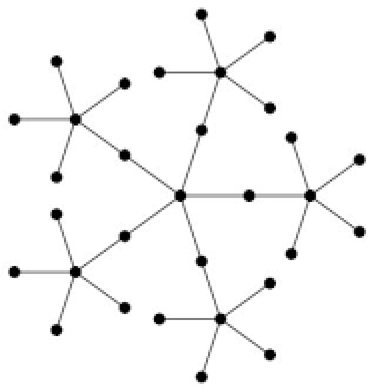




- Complete graph case:
  - Probably small number of comments
  - Could be isolated from the whole network
- Tree case:
  - People are so different within the chatting group that they do not share common topic

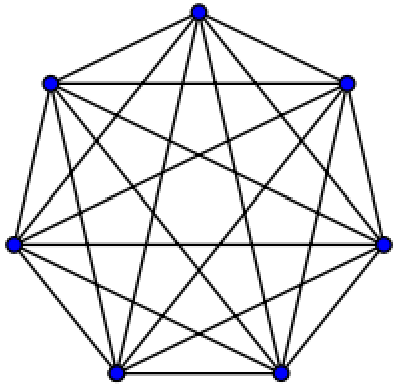
Extreme Case Discussion

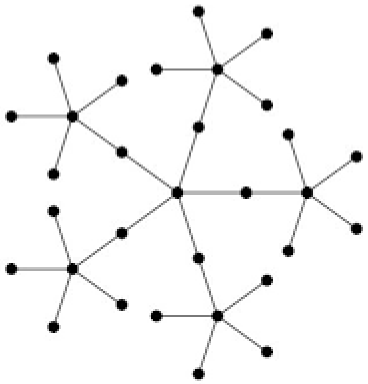




- Complete graph case:
  - Probably small number of comments
  - Could be isolated from the whole network
- Tree case:
  - People are so different within the chatting group that they do not share common topic

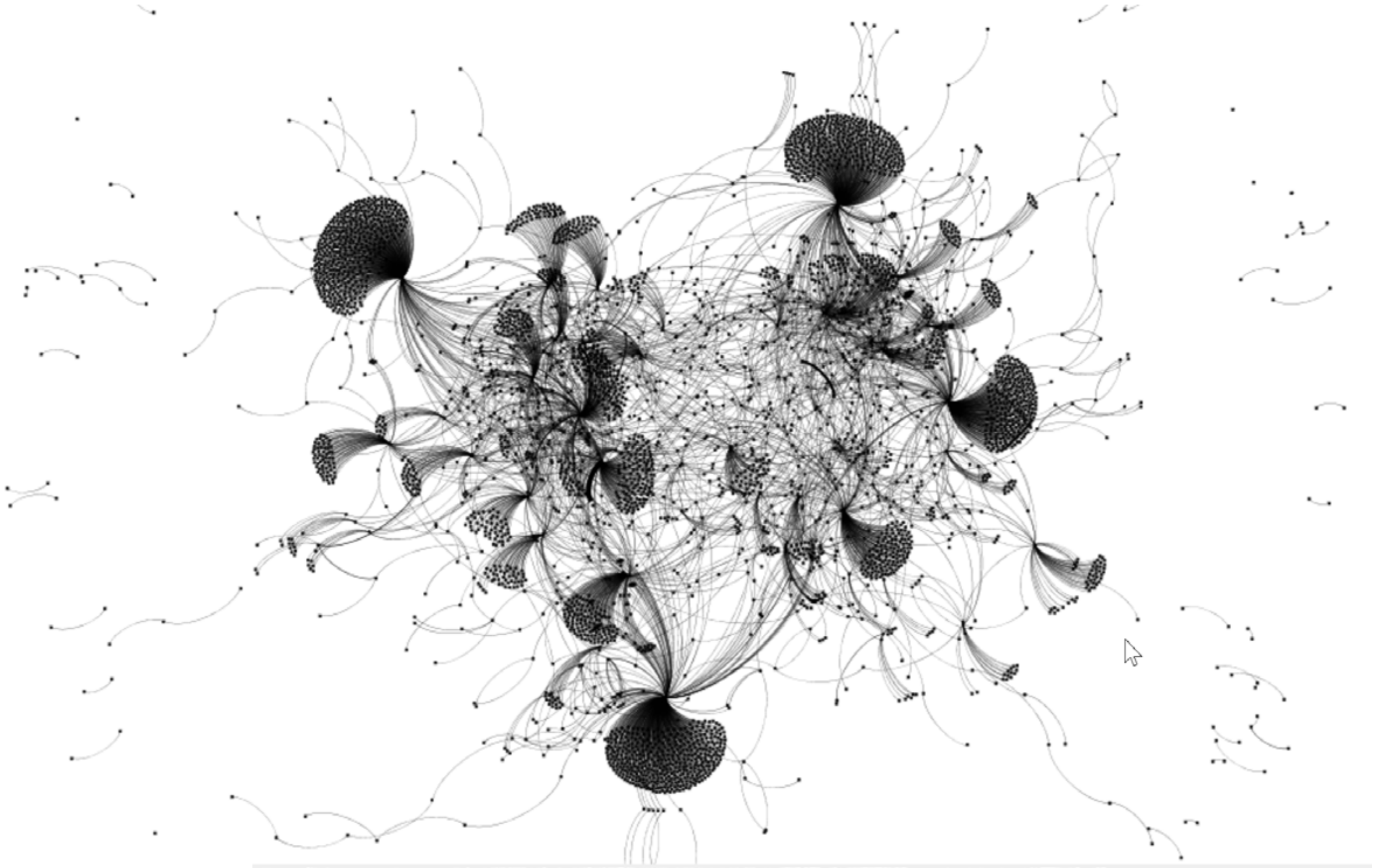
Extreme Case Discussion





- Complete graph case:
  - Probably small number of comments
  - Could be isolated from the whole network
- Tree case:
  - People are so different within the chatting group that they do not share common topic

Result-Subject A

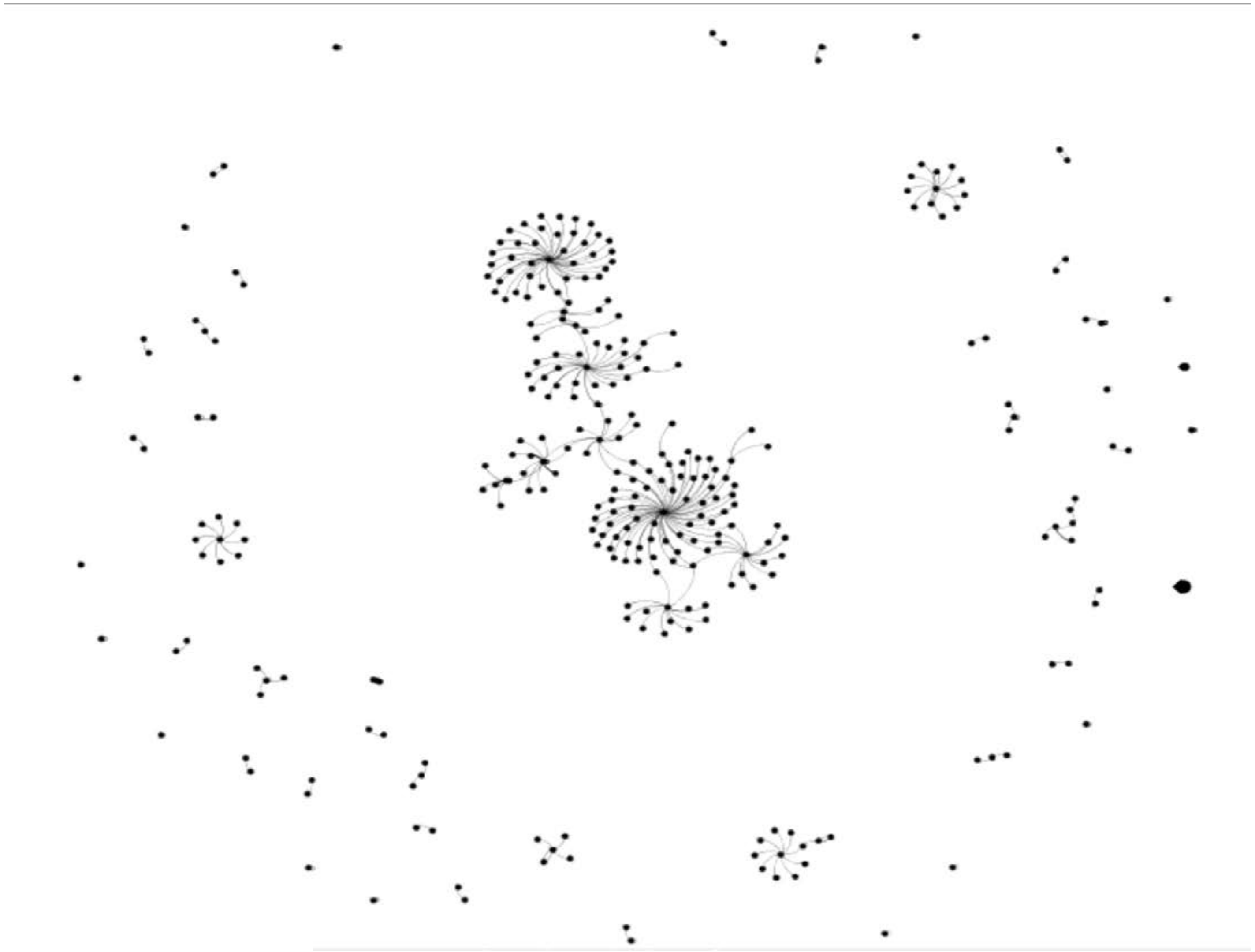


Number of comment: 266,916

Average Degree	1.26
Average Weighted Degree	1.453
Graph Density	0.001
Modularity	0.79
Clustering Coefficient	0.012
Average Path Length	5.507

Table 4

Result-Subject B

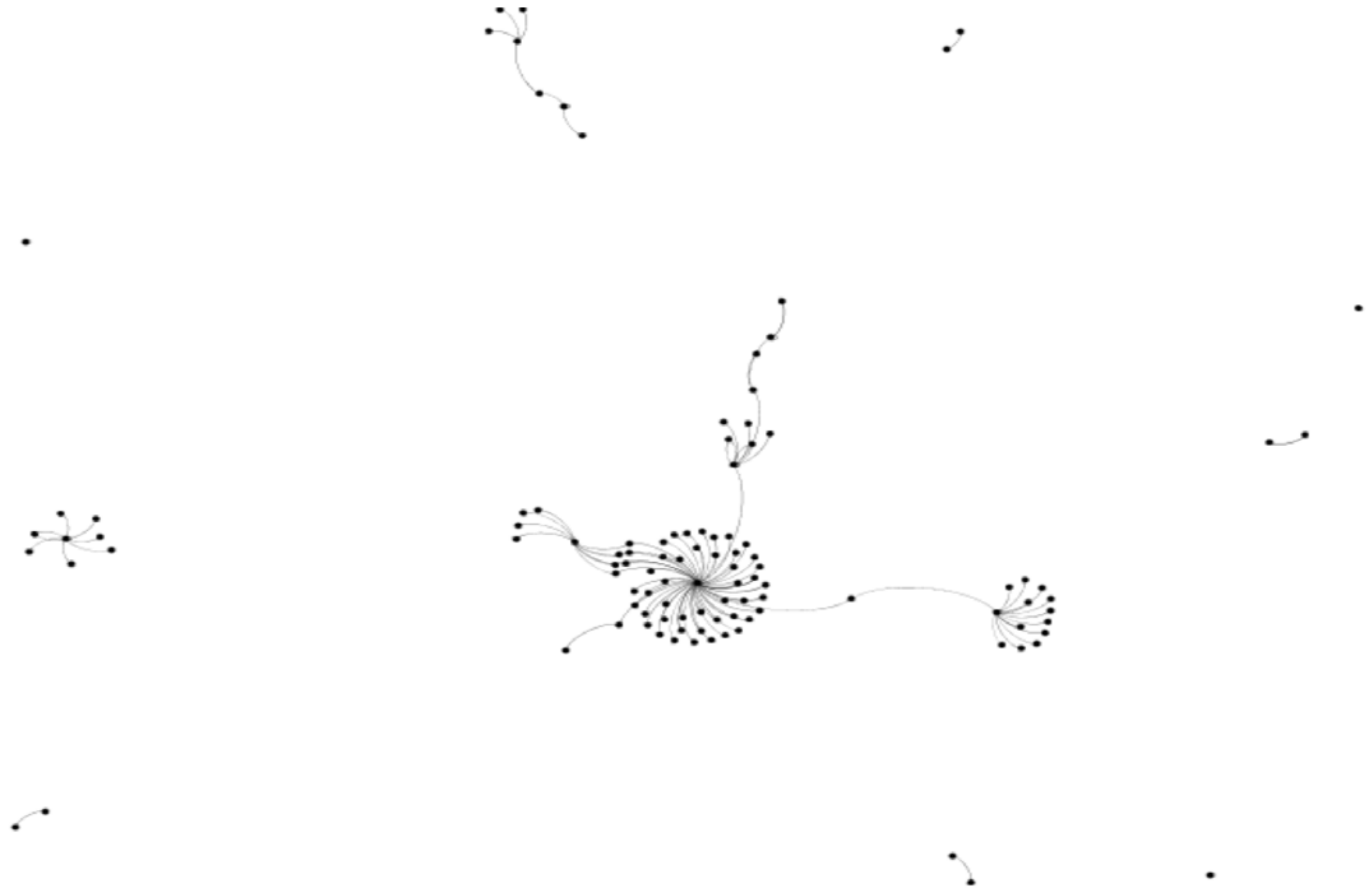


Number of comment: 102,199

Average Degree	1.006
Average Weighted Degree	1.392
Graph Density	0.003
Modularity	0.841
Clustering Coefficient	0.016
Average Path Length	2.052

Table 5

Result-Subject C



Number of comment: 10,981

Average Degree	1.054
Average Weighted Degree	1.241
Graph Density	0.009
Modularity	0.629
Clustering Coefficient	0.018
Average Path Length	1.176

Table 6

Primary Conclusion

- Co-movement between average path length and absolute number of comments
- Co-movement between modularity and relative number of comments

Primary Conclusion

- Co-movement between average path length and absolute number of comments
- Co-movement between modularity and relative number of comments

Why Useful

- Typology for analyzing fan groups
- Provide foundation for further public relation studies

Why Useful

- Typology for analyzing fan groups
- Provide foundation for further public relation studies

Speaker Transition

Does high-frequency trading eliminate behavioral biases?

Wanlin Ji

Background

- Market makers
- HFT on T-bond futures quoting

Background

- Market makers
- HFT on T-bond futures quoting

Background

- Market makers
- HFT on T-bond futures quoting

Background

- Market makers
- HFT on T-bond futures quoting

Background

- Market makers
- HFT on T-bond futures quoting

Theories

- disposition effect
- house-money effect
- self-attribution bias

Theories

- disposition effect
- house-money effect
- self-attribution bias

Market Order book

Royal Dutch Shell PLC RDSB xd Close 1,935 GBX

NMS	100,000	Segment	SET1	Sector	FF10	ISIN	GB00B03MM408
Period	OBT	Market	SETS				

Last 1,936.6o at 15:49:06 vol 65

Prev 1,937 AT 1,936 AT 1,937 AT 1,937 AT 1,937 AT 1,937 AT

Trade Hi	1,944	Open	1,925	Current	1,937
Trade Lo	1,919	VWAP	1,932.50	Current Hi	1,944
Total Vol	3,486,000	AT Vol	2,908,358 (83%)	Current Lo	1,919

Buy Vol	433,455			Sell Vol	506,918		
Buy Depth	49	VWAP	1,893.59	Sell Depth	87	VWAP	2,009.18

BUY	2	6,781	1,936 -	1,938	1,453	3	SELL
	15:48	5,781	1,936	1,938	57	15:44	
	15:48	1,000	1,936	1,938	800	15:46	
	15:39	6,998	1,935	1,938	596	15:48	
	15:49	1,000	1,935	1,939	1,000	15:48	
	15:49	5,760	1,935	1,940	567	15:33	
	15:49	5,000	1,935	1,940	2,500	15:34	
	15:36	6,487	1,934	1,940	6,557	15:41	
	15:44	3,000	1,934	1,940	5,000	15:48	
	15:45	100	1,934	1,941	3,852	15:47	
	15:49	1,000	1,934	1,941	10,000	15:48	
	15:08	3,900	1,933	1,941	4,351	15:48	
	15:36	6,598	1,933	1,942	2,667	15:22	
	15:39	11,000	1,933	1,943	2,766	15:15	
	15:44	4,000	1,933	1,943	5,760	15:41	
	15:44	7,139	1,932	1,943	500	15:48	
	15:46	10,000	1,932	1,945	10,000	08:35	

Market Order book

Royal Dutch Shell PLC RDSB xd Close 1,935 GBX

NMS 100,000 Segment SET1 Sector FF10 ISIN GB00B03MM408
Period OBT Market SETS

Last 1,936.6o at 15:49:06 vol 65
Prev 1,937 AT 1,936 AT 1,937 AT 1,937 AT 1,937 AT 1,937 AT

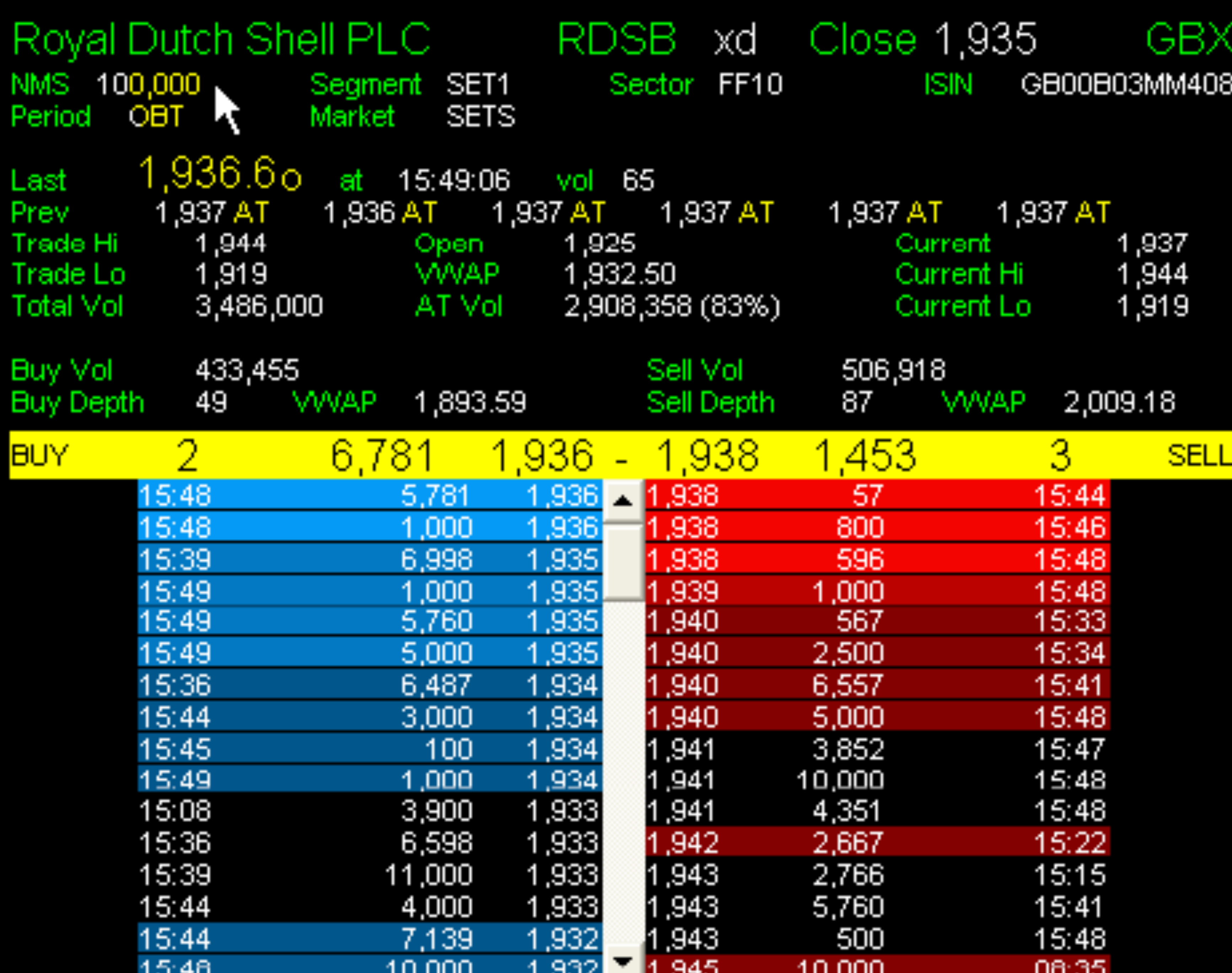
Trade Hi	1,944	Open	1,925	Current	1,937
Trade Lo	1,919	VWAP	1,932.50	Current Hi	1,944
Total Vol	3,486,000	AT Vol	2,908,358 (83%)	Current Lo	1,919

Buy Vol	433,455			Sell Vol	506,918		
Buy Depth	49	VWAP	1,893.59	Sell Depth	87	VWAP	2,009.18

BUY	2	6,781	1,936	-	1,938	1,453	3	SELL

15:48	5,781	1,936	1,938	57	15:44
15:48	1,000	1,936	1,938	800	15:46
15:39	6,998	1,935	1,938	596	15:48
15:49	1,000	1,935	1,939	1,000	15:48
15:49	5,760	1,935	1,940	567	15:33
15:49	5,000	1,935	1,940	2,500	15:34
15:36	6,487	1,934	1,940	6,557	15:41
15:44	3,000	1,934	1,940	5,000	15:48
15:45	100	1,934	1,941	3,852	15:47
15:49	1,000	1,934	1,941	10,000	15:48
15:08	3,900	1,933	1,941	4,351	15:48
15:36	6,598	1,933	1,942	2,667	15:22
15:39	11,000	1,933	1,943	2,766	15:15
15:44	4,000	1,933	1,943	5,760	15:41
15:44	7,139	1,932	1,943	500	15:48
15:46	10,000	1,932	1,945	10,000	08:35

Market Order book



Experimental Design

- Identifying 1-day horizon, micro-level
- Profits and risks
- Risk is measured in three different ways, as the number of afternoon trades, the average size of afternoon trades, or the cumulative risk-weighted inventory of each trader.
- Morning Losses Lead to Afternoon Risk-Taking

Experimental Design

- Identifying 1-day horizon, micro-level
- Profits and risks
- Risk is measured in three different ways, as the number of afternoon trades, the average size of afternoon trades, or the cumulative risk-weighted inventory of each trader.
- Morning Losses Lead to Afternoon Risk-Taking

Experimental Design

- Identifying 1-day horizon, micro-level
- Profits and risks
- Risk is measured in three different ways, as the number of afternoon trades, the average size of afternoon trades, or the cumulative risk-weighted inventory of each trader.
- Morning Losses Lead to Afternoon Risk-Taking

Experimental Design

- Identifying 1-day horizon, micro-level
- Profits and risks
- Risk is measured in three different ways, as the number of afternoon trades, the average size of afternoon trades, or the cumulative risk-weighted inventory of each trader.
- Morning Losses Lead to Afternoon Risk-Taking

Experimental Design

- Identifying 1-day horizon, micro-level
- Profits and risks
- Risk is measured in three different ways, as the number of afternoon trades, the average size of afternoon trades, or the cumulative risk-weighted inventory of each trader.
- Morning Losses Lead to Afternoon Risk-Taking

Panel A: Statistics by Trader-Day

Variable	Morning Mean	Morning Median	Morning St. Dev.	Afternoon Mean	Afternoon Median	Afternoon St. Dev.
All Trader-Days ($N = 82{,}595$) Raw Data						
Profits	1808.33	750.00	171848.13	661.78	187.50	113964.28
Number of trades	116.62	88.00	105.37	73.25	52.00	72.95
Average trade size	10.03	4.84	19.17	9.35	4.53	18.27
Total dollar risk	9641.46	1150.00	57540.27	10876.76	1242.83	75133.82
Price-setting trades	0.202	0.000	0.514	0.327	0.000	0.643
Traders with Profitable Mornings ($N = 55{,}877$) Normalized by Trader						
Profits	0.467	0.276	0.574	0.095	0.067	0.733
Number of trades	−0.035	−0.159	0.986	−0.066	−0.234	0.980
Average trade size	−0.063	−0.222	0.967	−0.046	−0.213	0.989
Total dollar risk	−0.122	−0.317	0.776	−0.100	−0.335	0.801
Price-setting trades	−0.009	−0.188	0.601	−0.017	−0.128	0.467
Traders with Losing Mornings ($N = 26{,}718$) Normalized by Trader						
Profits	−0.563	−0.273	0.727	0.082	0.067	0.915
Number of trades	0.066	−0.065	1.013	0.124	−0.036	1.016
Average trade size	0.119	−0.081	1.040	0.086	−0.114	1.006
Total dollar risk	0.180	−0.146	0.993	0.141	−0.205	0.997
Price-setting trades	0.018	−0.171	0.619	0.036	−0.116	0.526

Panel B: Statistics by Day

Variable	Mean	St. Dev.	Minimum	Maximum
Afternoon price changes	621.8703	215.383	195.00	1582.00
Fraction with morning losses	0.3238	0.049	0.20	0.50
Fraction of loss-averse traders with losses	0.3305	0.055	0.19	0.50
Fraction of price-setting traders with losses	0.3230	0.051	0.19	0.49

Panel A: Statistics by Trader-Day						
	Morning			Afternoon		
Variable	Mean	Median	St. Dev.	Mean	Median	St. Dev.
All Trader-Days ($N = 82{,}595$) Raw Data						
Profits	1808.33	750.00	171848.13	661.78	187.50	113964.28
Number of trades	116.62	88.00	105.37	73.25	52.00	72.95
Average trade size	10.03	4.84	19.17	9.35	4.53	18.27
Total dollar risk	9641.46	1150.00	57540.27	10876.76	1242.83	75133.82
Price-setting trades	0.202	0.000	0.514	0.327	0.000	0.643
Traders with Profitable Mornings ($N = 55{,}877$) Normalized by Trader						
Profits	0.467	0.276	0.574	0.095	0.067	0.733
Number of trades	−0.035	−0.159	0.986	−0.066	−0.234	0.980
Average trade size	−0.063	−0.222	0.967	−0.046	−0.213	0.989
Total dollar risk	−0.122	−0.317	0.776	−0.100	−0.335	0.801
Price-setting trades	−0.009	−0.188	0.601	−0.017	−0.128	0.467
Traders with Losing Mornings ($N = 26{,}718$) Normalized by Trader						
Profits	−0.563	−0.273	0.727	0.082	0.067	0.915
Number of trades	0.066	−0.065	1.013	0.124	−0.036	1.016
Average trade size	0.119	−0.081	1.040	0.086	−0.114	1.006
Total dollar risk	0.180	−0.146	0.993	0.141	−0.205	0.997
Price-setting trades	0.018	−0.171	0.619	0.036	−0.116	0.526

Panel B: Statistics by Day				
Variable	Mean	St. Dev.	Minimum	Maximum
Afternoon price changes	621.8703	215.383	195.00	1582.00
Fraction with morning losses	0.3238	0.049	0.20	0.50
Fraction of loss-averse traders with losses	0.3305	0.055	0.19	0.50
Fraction of price-setting traders with losses	0.3230	0.051	0.19	0.49

Panel A: Statistics by Trader-Day						
	Morning			Afternoon		
Variable	Mean	Median	St. Dev.	Mean	Median	St. Dev.
All Trader-Days (N = 82,595) Raw Data						
Profits	1808.33	750.00	171848.13	661.78	187.50	113964.28
Number of trades	116.62	88.00	105.37	73.25	52.00	72.95
Average trade size	10.03	4.84	19.17	9.35	4.53	18.27
Total dollar risk	9641.46	1150.00	57540.27	10876.76	1242.83	75133.82
Price-setting trades	0.202	0.000	0.514	0.327	0.000	0.643
Traders with Profitable Mornings (N = 55,877) Normalized by Trader						
Profits	0.467	0.276	0.574	0.095	0.067	0.733
Number of trades	−0.035	−0.159	0.986	−0.066	−0.234	0.980
Average trade size	−0.063	−0.222	0.967	−0.046	−0.213	0.989
Total dollar risk	−0.122	−0.317	0.776	−0.100	−0.335	0.801
Price-setting trades	−0.009	−0.188	0.601	−0.017	−0.128	0.467
Traders with Losing Mornings (N = 26,718) Normalized by Trader						
Profits	−0.563	−0.273	0.727	0.082	0.067	0.915
Number of trades	0.066	−0.065	1.013	0.124	−0.036	1.016
Average trade size	0.119	−0.081	1.040	0.086	−0.114	1.006
Total dollar risk	0.180	−0.146	0.993	0.141	−0.205	0.997
Price-setting trades	0.018	−0.171	0.619	0.036	−0.116	0.526

Panel B: Statistics by Day				
Variable	Mean	St. Dev.	Minimum	Maximum
Afternoon price changes	621.8703	215.383	195.00	1582.00
Fraction with morning losses	0.3238	0.049	0.20	0.50
Fraction of loss-averse traders with losses	0.3305	0.055	0.19	0.50
Fraction of price-setting traders with losses	0.3230	0.051	0.19	0.49

Reference

- Bian, Jiangze, et al. ”Do Behavioral Biases Affect Order Aggressiveness?.”, Review of Finance (2017): rfx037
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. ”High-frequency trading and price discovery.” The Review of Financial Studies 27.8 (2014): 2267- 2306.
- Budish, Eric, Peter Cramton, and John Shim. ”The high-frequency trading arms race: Frequent batch auctions as a market design response.” The Quarterly Journal of Economics 130.4 (2015): 1547-1621.
- Coval, Joshua D., and Tyler Shumway. ”Do behavioral biases affect prices?.” The Journal of Finance 60.1 (2005): 1-34.
- Davis, Ryan L., Bonnie F. Van Ness, and Robert A. Van Ness. ”Clustering of Trade Prices by HighFrequency and NonHighFrequency Trading Firms.” Financial Review 49.2 (2014): 421-433.
- Li, Kun, Rick Cooper, and Ben Van Vliet. ”How Does High-Frequency Trading Affect Low-Frequency Trading?.” Journal of Behavioral Finance (2017): 1-14.

Speaker Transition

Impact of World Bank's projects performance:

Country level indicators vs. project level indicators

Zhuo Leng

March, 2018

Outline

1. Problem Formulation
2. Research Question
3. Introduction
4. Literature Review
5. Data
6. Method
7. Result
8. Conclusion and future work

Problem Formulation

> “projects with high monitoring and evaluation quality score between 0.13 and 0.40 points better than projects with poor M&E quality on a six-point outcome scale”

Raimondo (2016). What Difference Does Good Monitoring & Evaluation Make to World Bank Project Performance?

Problem Formulation

- Problem Formulation
  - Each project is evaluated by the Independent Evaluation Group after its completion which is not time-efficient
  - For the already-happened bad performance projects, there is nothing can be improved
- Goal
  - Make the performance prediction of the project during the project approval period

Research Question

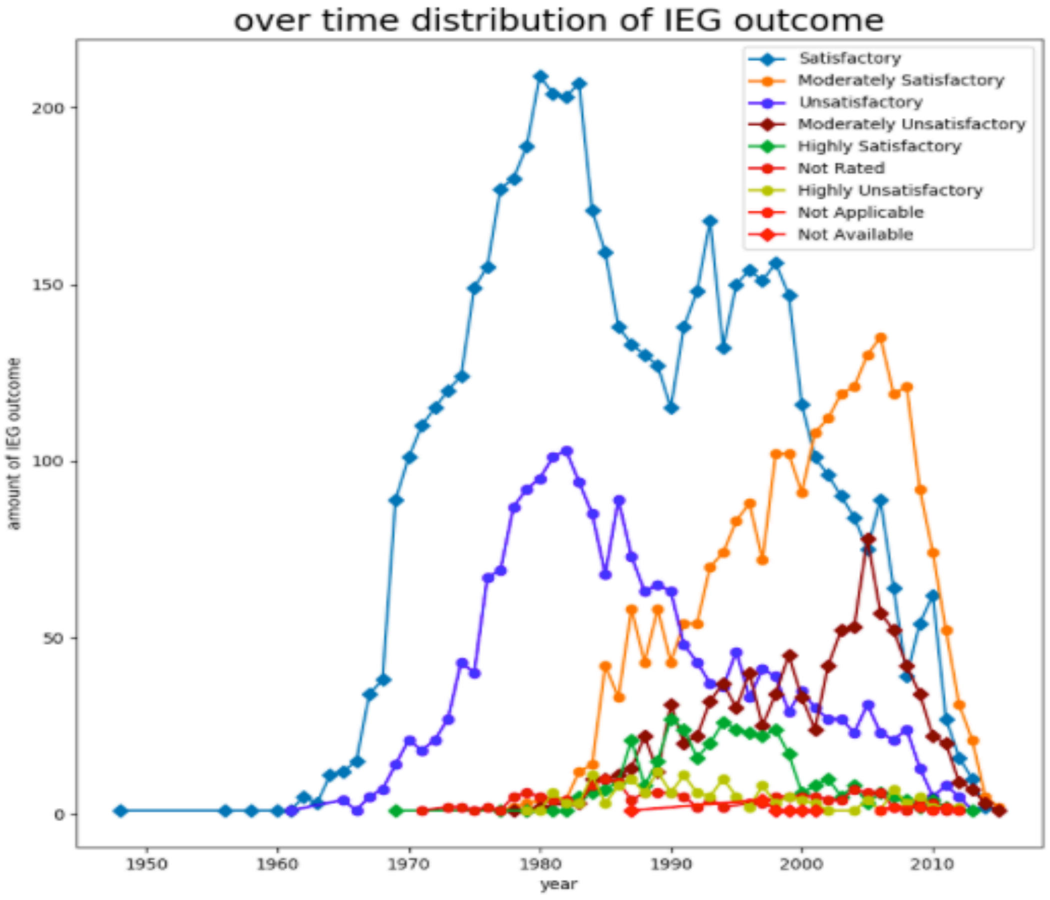
Research Question

My research is going to investigates how country level indicators (e.g. GDP, Political Stability, government debt (total % of GDP), etc.) and project level measures (e.g.project duration, countries, etc.) affect the World Bank project assessment. Then Make the performance prediction of the project during the project approval period

Data Description

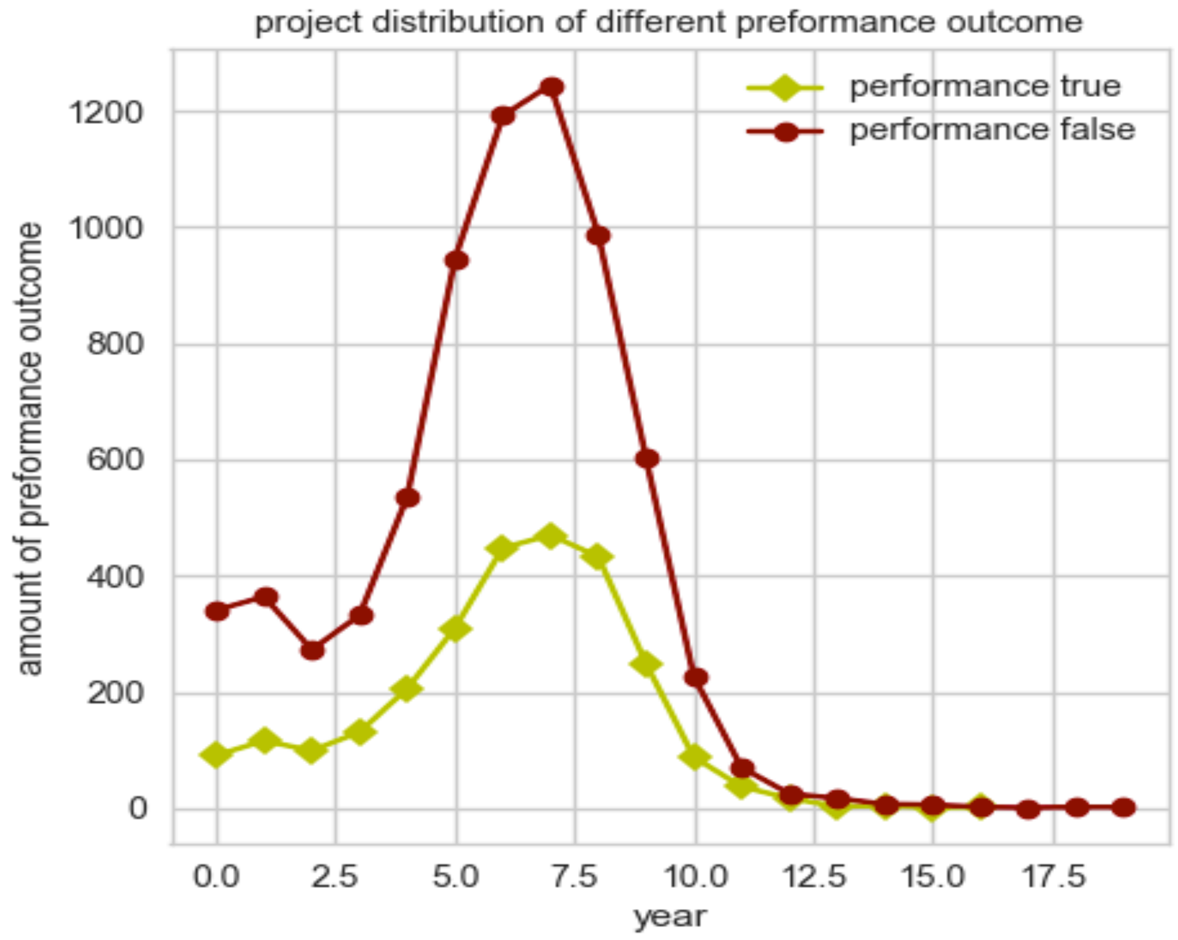
World Bank Projects API	*World Bank Project Performance Ratings*	*World Economic Development Indicators*	*Worldwide Governance Indicators*
Detailed information about all WB projects	Assessment of a sample of projects, by the Independent Evaluation Group	The most current and accurate global development data, includes national, regional and global estimates.	Include individual governance indicators for over 200 countries and territories over the period 1996-2015.

Data Description



Feature Type	Features Selected
Numerical	leading project cost, ibrd commitment amount, ida commitment amount, total commitment amount, life expectancy at birth, the GDP index,central government debt total (% of GDP), other governance indicators...
Categorical	region, country name, product line, leading instr type, agreement type
Aggregated	total number of projects done for the past years by the each country, total commitment amount for the past years, average GDP, average life span index, cost-commitment ratio, project duration time

3. Data Description



- No statistical difference in project duration between Unsatisfactory (“true”) and Satisfactory projects
- 11726 projects with the approval year from 1948 to 2015, most project duration is 6-7.5 years
- Concerns: depending on the long duration of the projects, the information we get at the starting point may be not enough or valuable

Project performance, summary information

- Total projects in dataset: 9,884

- “Unsatisfactory” projects: 27.3%

- Training/testing: 1967 to 2012

Research Method

- Multinomial Logistic Regression Model vs. Logistic Regression Model

$(y = 1) = \log p(y = 1) = \beta_0 + \beta_1 \text{feature1} + \beta_2 \text{feature2} + \ldots \beta_{12} \text{featuren}$

- Random Forest

- Neural Network

Data-Preprocessing

- Outlier/missing value treatment
- Check correlation matrix
- Create two types of dependent variable set:
  - 6 categories: Satisfactory, Moderately Satisfactory, Unsatisfactory, Moderately Unsatisfactory, Highly Satisfactory, Highly Unsatisfactory,

  - 2 categories:
    - Satisfactory: Satisfactory, Moderately Satisfactory, , Highly Satisfactory:
    - Unsatisfactory: Unsatisfactory, Moderately Unsatisfactory, Highly Unsatisfactory

- Split data into training/testing set:7:3

Pairwise Correlation Matrix

lendinginstrtype_AD	lendinginstrtype_IN	0.998796
lendinginstrtype_IN	lendinginstrtype_AD	0.998796
sum_amt_proj	total_num_proj	0.947389
total_num_proj	sum_amt_proj	0.947389
totalcommamt	ibrdcommamt	0.929069
ibrdcommamt	totalcommamt	0.929069
Agreement Type_IBRD	Agreement Type_IDA	0.915903
Agreement Type_IDA	Agreement Type_IBRD	0.915903
avg_lifespan	SP.DYN.LE00.IN	0.893750
SP.DYN.LE00.IN	avg_lifespan	0.893750
Region_AFR	avg_lifespan	0.788593
avg_lifespan	Region_AFR	0.788593
avg_GDP	NY.GDP.PCAP.CD	0.762048
NY.GDP.PCAP.CD	avg_GDP	0.762048
SP.DYN.LE00.IN	Region_AFR	0.704933
Region_AFR	SP.DYN.LE00.IN	0.704933
Product Line_IBRD/IDA	Agreement Type_GEF	0.681009
Agreement Type_GEF	Product Line_IBRD/IDA	0.681009
Product Line_IBRD/IDA	Product Line_Global Environment Project	0.681009
Region_MNA	Country Name_Morocco	0.547435
Country Name_Morocco	Region_MNA	0.547435
avg_GDP	Agreement Type_IDA	0.519822
Agreement Type_IDA	avg_GDP	0.519822
lendinginstr_Structural Adjustment Loan	lendinginstrtype_AD	0.517571
lendinginstrtype_AD	lendinginstr_Structural Adjustment Loan	0.517571
lendinginstrtype_IN	lendinginstr_Structural Adjustment Loan	0.516948
lendinginstr_Structural Adjustment Loan	lendinginstrtype_IN	0.516948
Region_LCR	avg_GDP	0.512896
avg_GDP	Region_LCR	0.512896
Region_MNA	Country Name_Tunisia	0.509500
Country Name_Tunisia	Region_MNA	0.509500
lendprojectcost	amt_ratio	0.507962
amt_ratio	lendprojectcost	0.507962
avg_GDP	Agreement Type_IBRD	0.501520
Agreement Type_IBRD	avg_GDP	0.501520

- By transfer all categorical features to dummy variables, now I got 178 features in total.
- Reduce the number of feature variables by looking at a pairwise correlation matrix. Cut out highly correlated variables

Multinomial Logistic Regression Model vs. Logistic Regression Model

Accuracy score on test set: 0.512 vs. 0.746

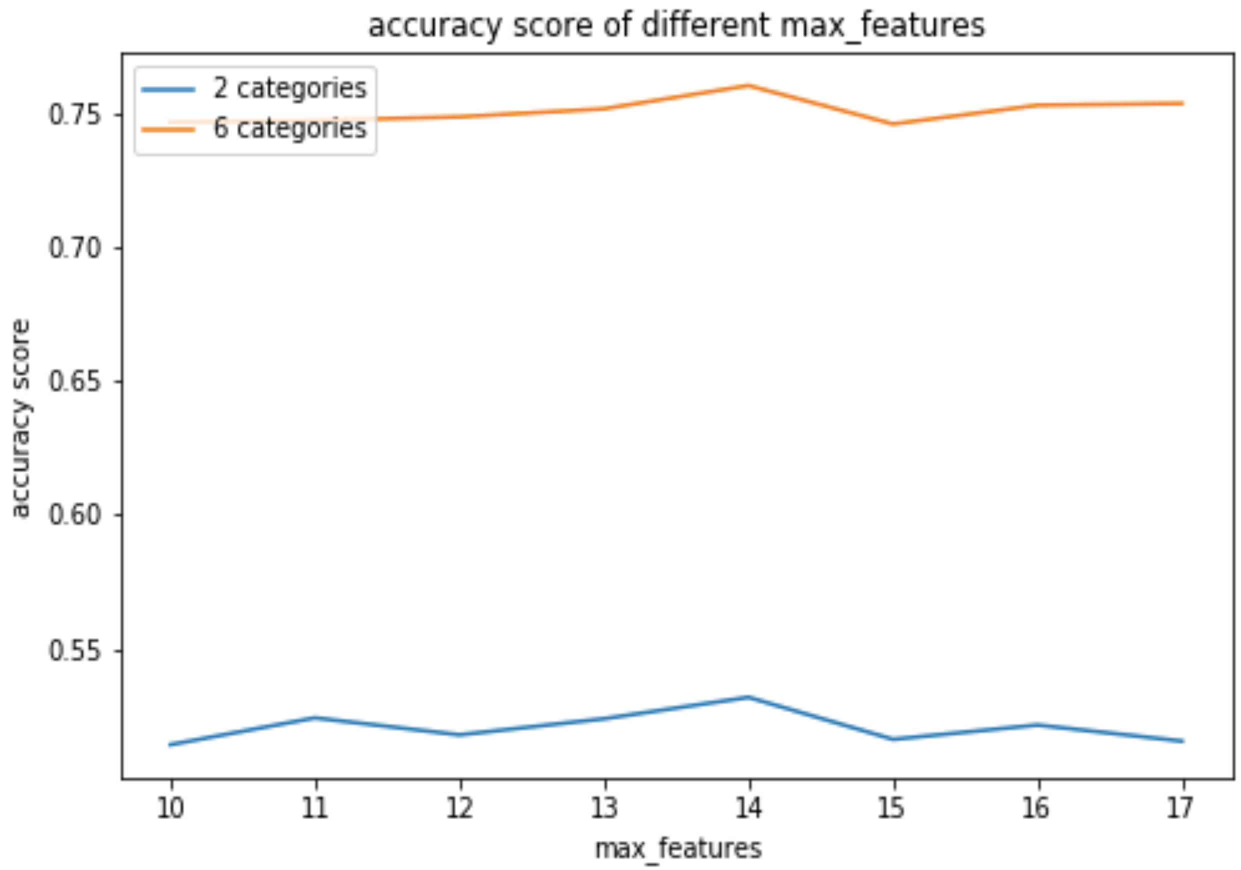
Significant features:

- Lendprojectcost
- Idacommamt
- proj_duration
- sum_amt_proj
- avg_GDP

Coefficient of significant features:

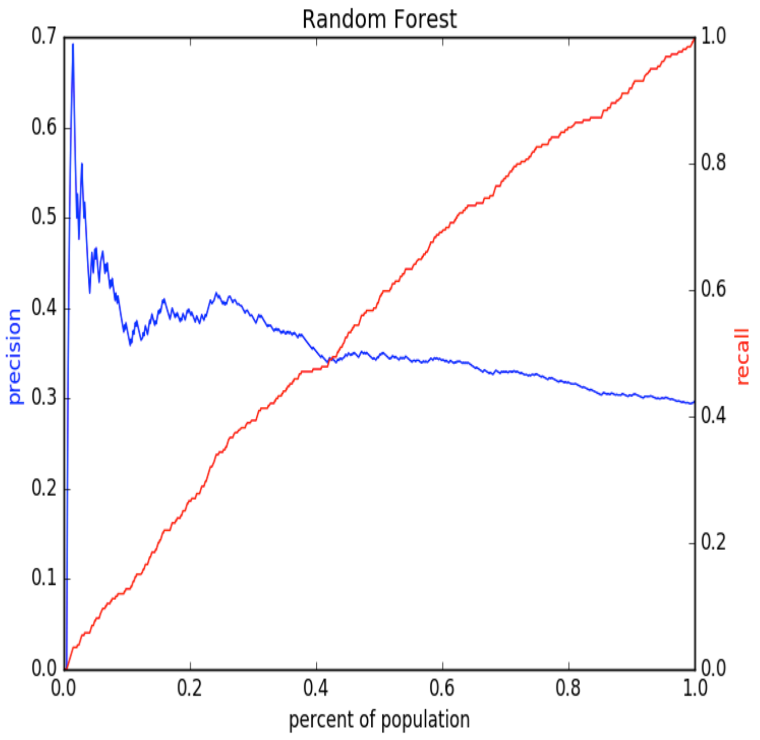
```
[-5.02908030e-09 -1.43565568e-12  9.75586738e-18 -3.83773683e-18
  8.34017207e-18]
```

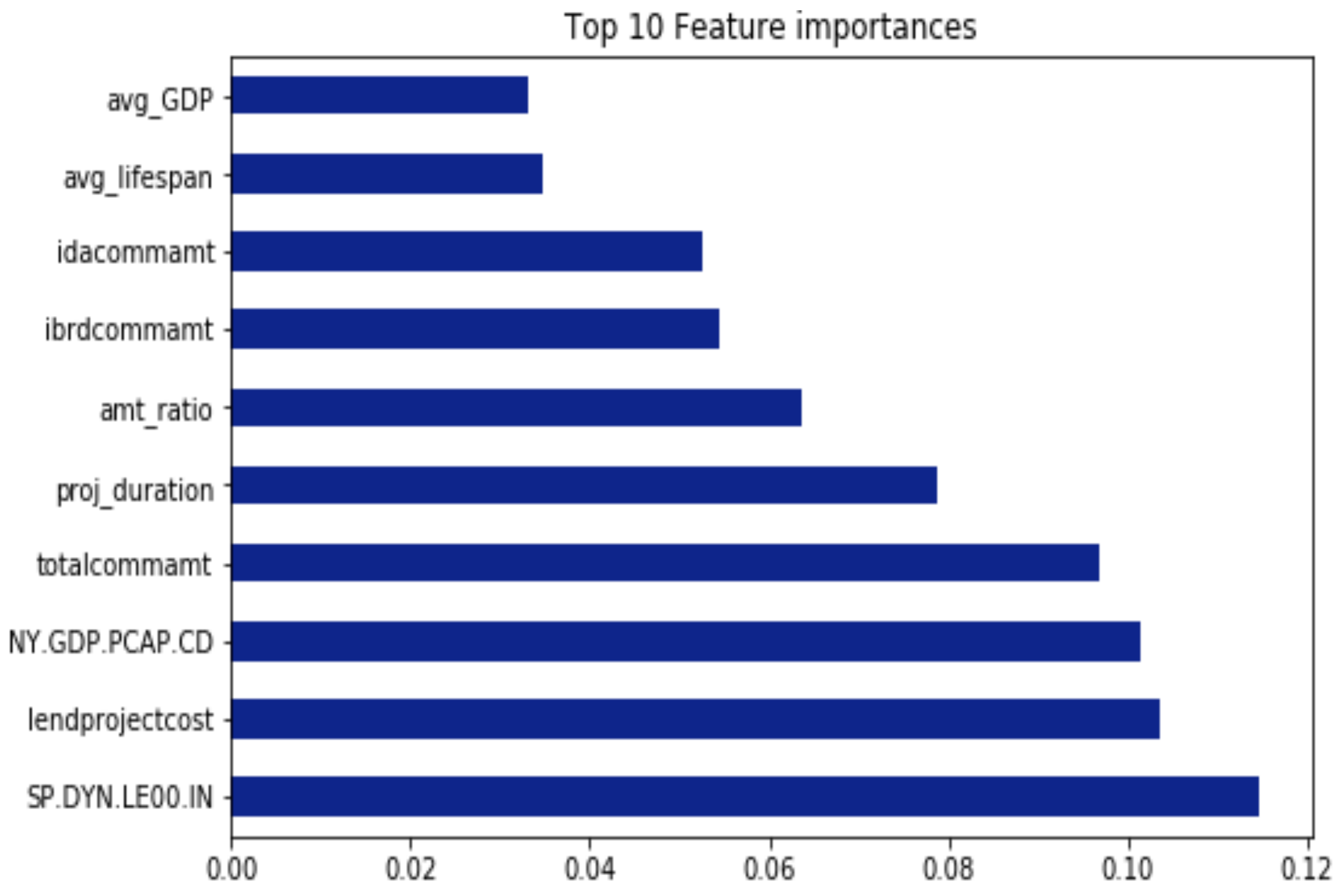
Random Forest



- Accuracy Score on test set with 6= 0.52153, which 2 categories = 0.76026
- Tune parameter: max_features. The best choice of max_features is close to square root of total features.

Random Forest





Neural Network

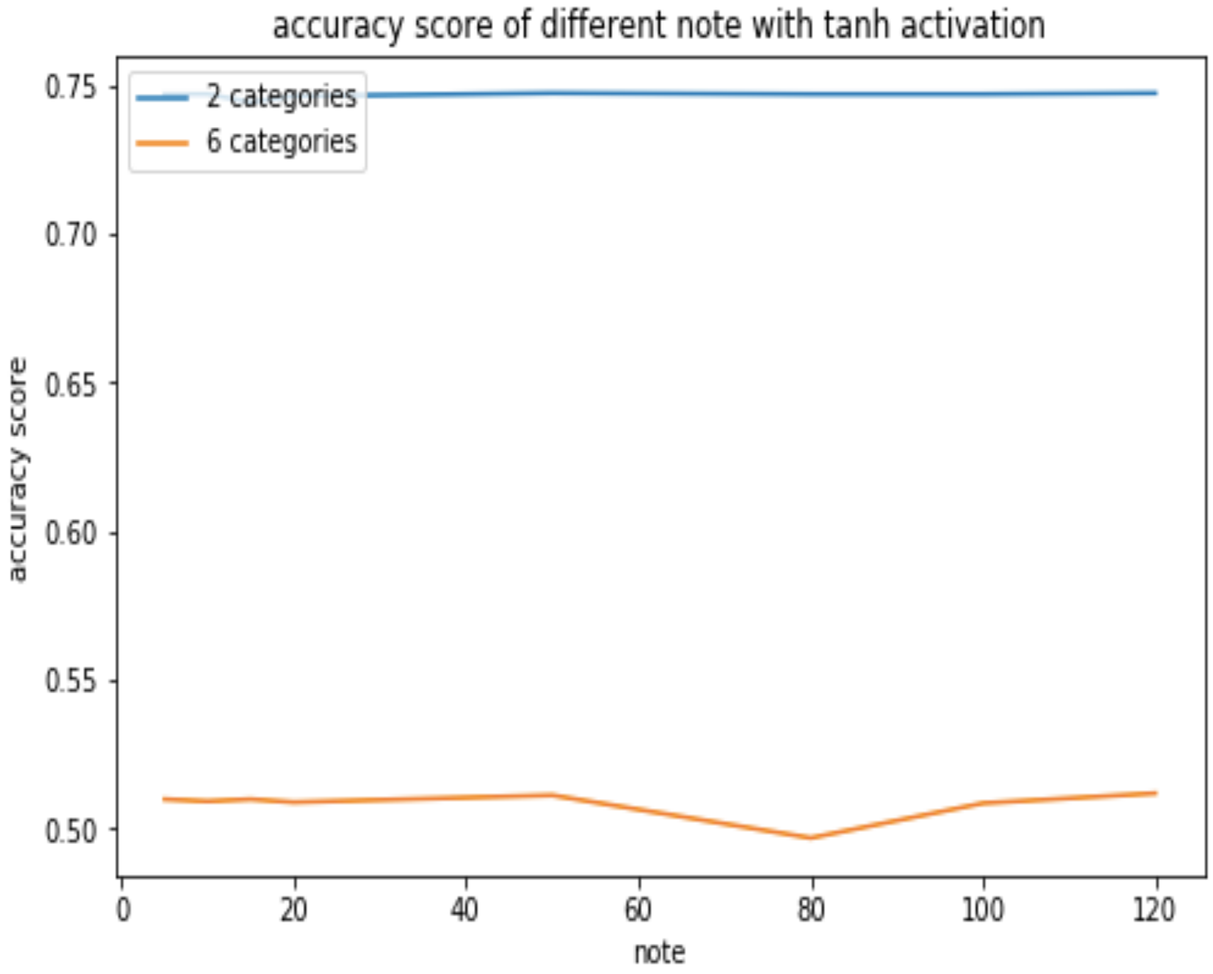
```
activation = ['identity','logistic', 'tanh', 'relu']
note = [5,10,15,20,50,80,100,120]
```

- Accuracy Score with 6= 0.51018,

```
['logistic', 80, 0.5101836393989984]
```

- with 2 categories = 0.746911

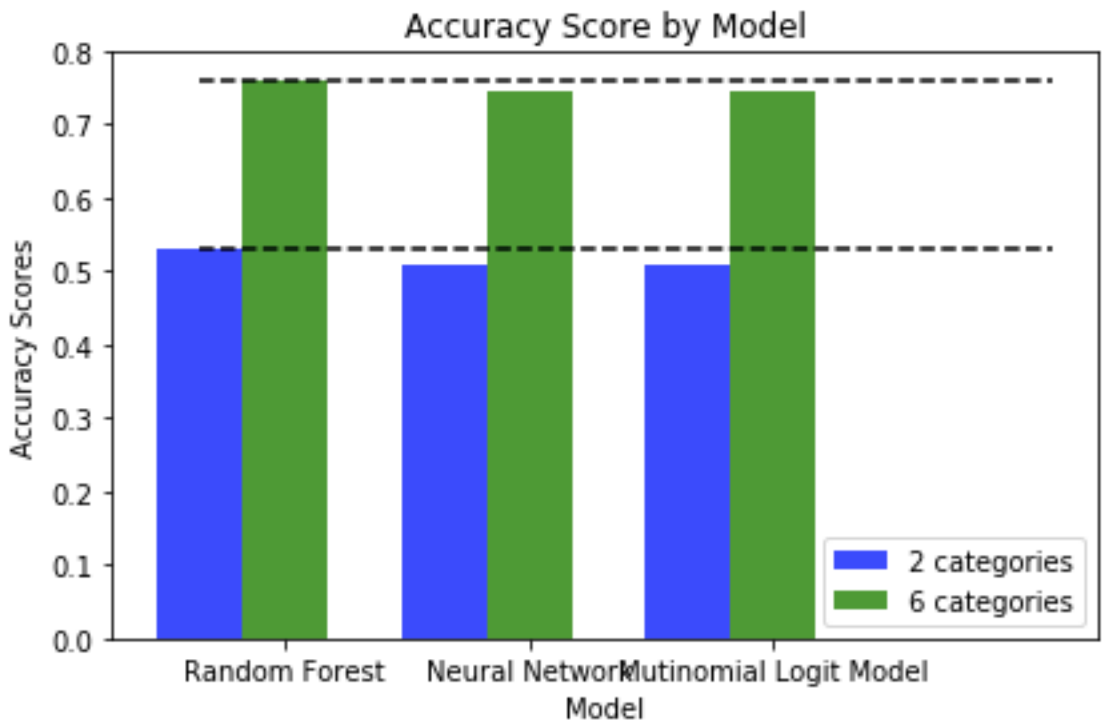
```
['identity', 15, 0.7469115191986644]
```



Conclusion

Model Selection

- Random Forest has the best performance



Conclusion

Initial Features Analysis/Selection

- Country characteristics (GDP, life expectancy) seem to have more predictive power then project level
- Country dummies yielded worse results (too few projects in some cases)
- Product Line and Agreement Type dummies seemed to generate bad results as well
- Governance characteristics not useful, too many missings

Future Work

- Create new features of country level
- Conduct PCA on Country level dummies to reduce categorical features' dimension
- Feature extraction using text analysis

THANKS

Zhuo Leng

Speaker Transition

Self-Affirmation and Opinion Malleability on r/ChangeMyView

Julian McClellan

What's Up

CMV: Nations whose leadership is based upon religion are fundamentally backwards and have no place in the modern world.

7 days ago by * (last edited 6 days ago) Anorak_

The separation of church and state has been around in the U.S. since 1802, when then-president Thomas Jefferson wrote the Danbury Papers assuring that the First Amendment did, in fact, ensure that the church and the state would exist as separate entities. Nations that have not adopted the same ideals and whose leadership is rooted in religion seem to generally lag behind economically, socially, and technologically in comparison.

Examples of this include: Turkey (more economically and socially than technologically), North Korea, China, Saudi Arabia, Qatar, and the majority of African religious nations (I realize that their lagging behind the rest of the world isn't solely due to leadership, but the constant political turmoil and tension often resulting from the religion-based political rule prevents them from making any meaningful progress)

What’s Up

CMV: Nations whose leadership is based upon religion are fundamentally backwards and have no place in the modern world.

7 days ago by * (last edited 6 days ago) Anorak_

The separation of church and state has been around in the U.S. since 1802, when then-president Thomas Jefferson wrote the Danbury Papers assuring that the First Amendment did, in fact, ensure that the church and the state would exist as separate entities. Nations that have not adopted the same ideals and whose leadership is rooted in religion seem to generally lag behind economically, socially, and technologically in comparison.

Examples of this include: Turkey (more economically and socially than technologically), North Korea, China, Saudi Arabia, Qatar, and the majority of African religious nations (I realize that their lagging behind the rest of the world isn't solely due to leadership, but the constant political turmoil and tension often resulting from the religion-based political rule prevents them from making any meaningful progress)

Self Affirmation Theory

“[I]f global self-worth is temporarily bolstered by success in a second, unrelated domain, the individual should be more willing to tolerate a threat to the domain of interest.”

--Correll (2004)

Self Affirmation Theory

“[I]f global self-worth is temporarily bolstered by success in a second, unrelated domain, the individual should be more willing to tolerate a threat to the domain of interest.”

--Correll (2004)

Self Affirmation Theory

- Only really been tested under lab conditions
  - Treatment is for subjects to write a bit about some held belief.
  - Then show them material opposing another of theirs
- Reddit user histories make it possible to look further into the past, to perhaps glimpse the extent to which self-affirmation theory applies.

Data

- All CMV Submissions Made in 2016
  - All comments (attempts to change views, deltas given, moderator comments, confirmation of deltas).
- All Submissions from Redditors (observation subjects) who made CMV Submissions in 2016
  - No comments

Data

- All CMV Submissions Made in 2016
  - All comments (attempts to change views, deltas given, moderator comments, confirmation of deltas).
- All Submissions from Redditors (observation subjects) who made CMV Submissions in 2016
  - No comments

Missing Data

- Submissions can be “[removed]” or “[deleted]”.
    - Non-random missing data issue

Missing Data

- Submissions can be “[removed]” or “[deleted]”.
  - Non-random missing data issue

Methods

- Binary Classification for the observation subjects' first CMV Submission with special interest in pre-debate variables but controlling for obvious post-debate variables as well.
- Pre-debatevariables include:
  - Number of previous submissions
  - Similarity Score of the content of the first CMV Submission to the author's corpus of previous submissions
- Post-debate:
  - Number of comments (direct, total, from author)
  - Salt

Methods

- Binary Classification for the observation subjects' first CMV Submission with special interest in pre-debate variables but controlling for obvious post-debate variables as well.
- Pre-debatevariables include:
  - Number of previous submissions
  - Similarity Score of the content of the first CMV Submission to the author's corpus of previous submissions
- Post-debate:
  - Number of comments (direct, total, from author)
  - Salt

Methods

- Binary Classification for the observation subjects' first CMV Submission with special interest in pre-debate variables but controlling for obvious post-debate variables as well.
- Pre-debate variables include:
  - Number of previous submissions
  - Similarity Score of the content of the first CMV Submission to the author's corpus of previous submissions
- Post-debate:
  - Number of comments (direct, total, from author)
  - Salt

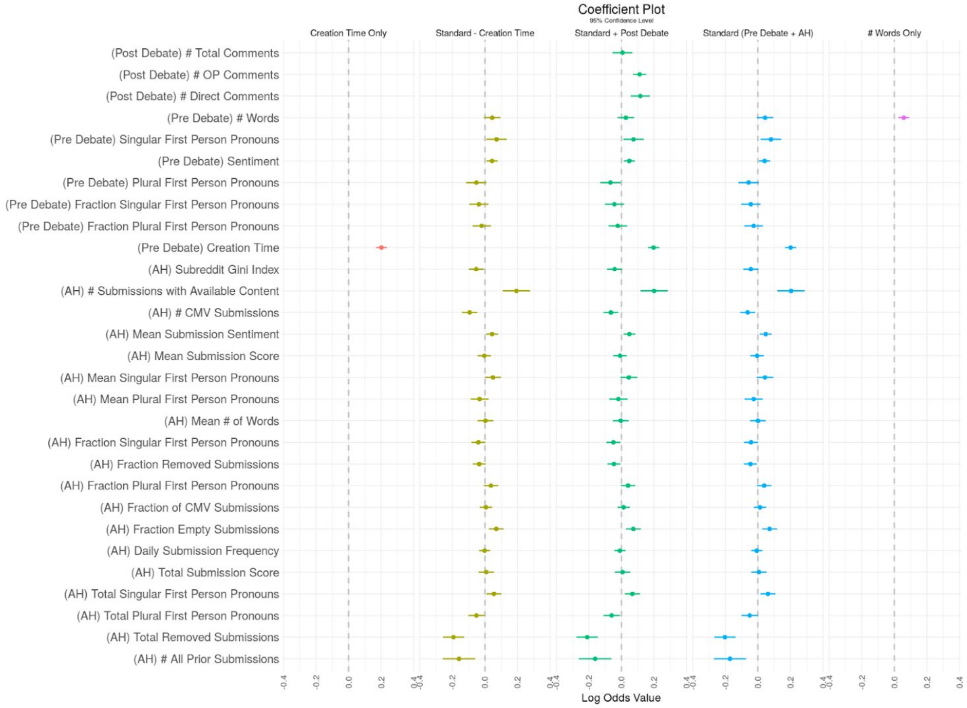
Similarity Score Methods

- Jaccard similarity
- Cosine Similarity
  - +Tf-idf
  - +Latent Semantic Analysis
    - Gives nice normal distribution and negative (dissimilar) scores

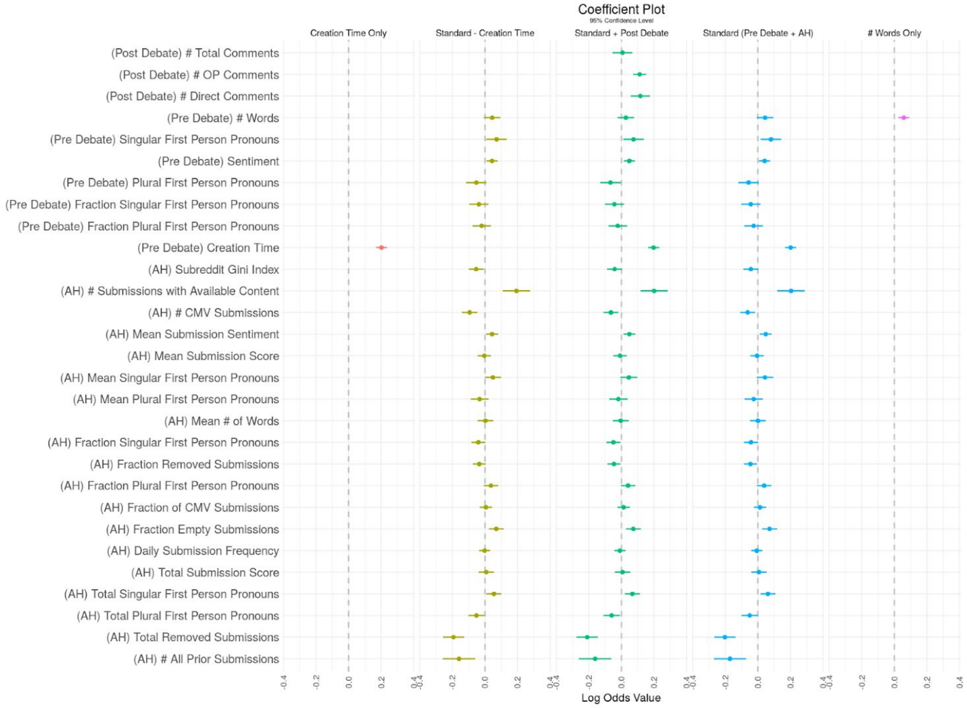
Similarity Score Methods

- Jaccard similarity
- Cosine Similarity
  - +Tf-idf
  - +Latent Semantic Analysis
    - Gives nice normal distribution and negative (dissimilar) scores

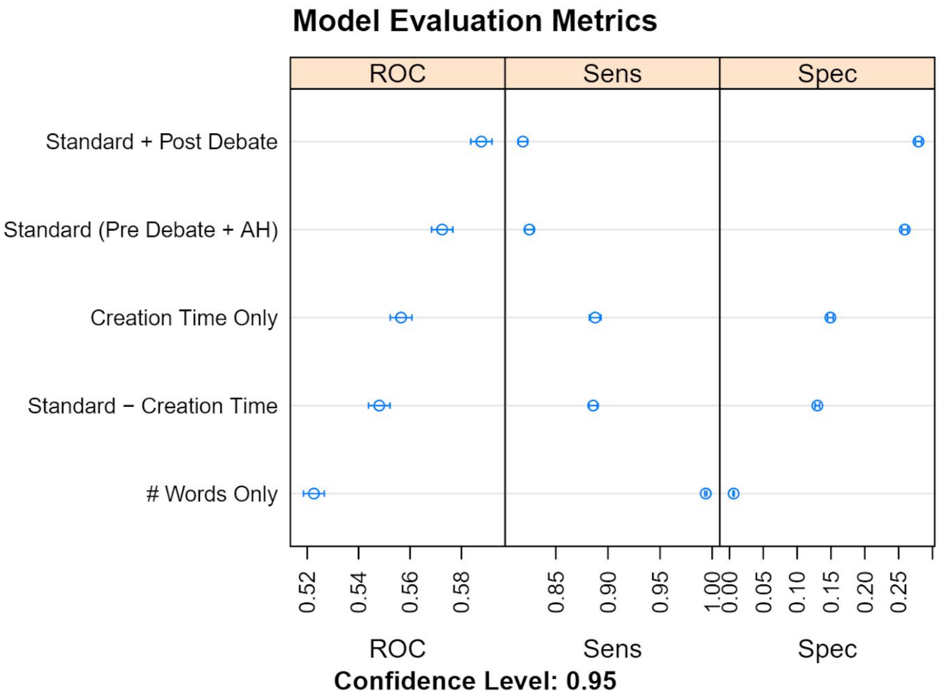
Expectations



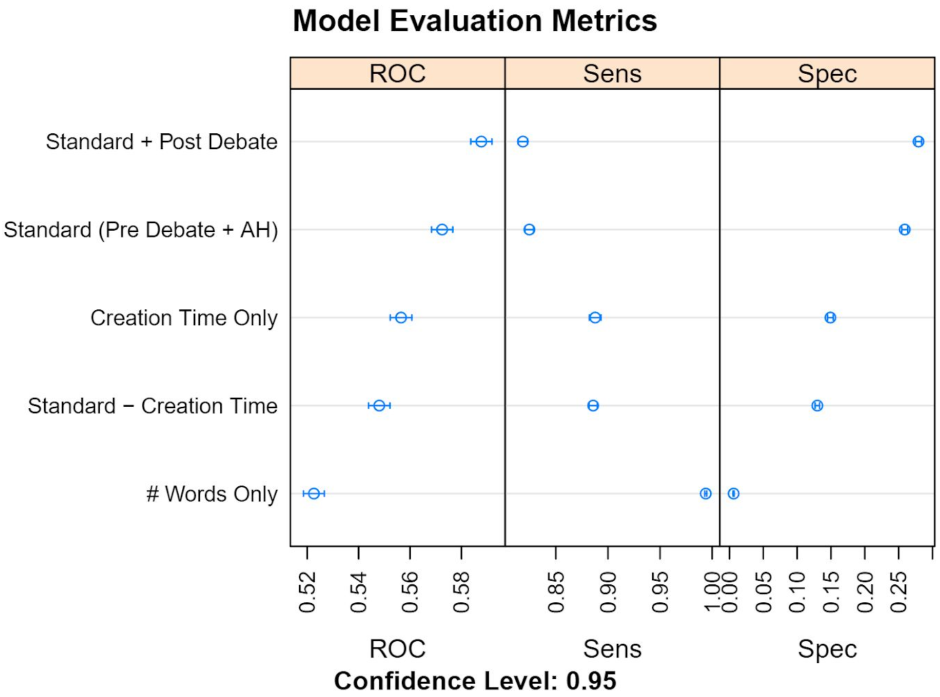
Expectations



Expectations



Expectations



15 Seconds Left

- How long does the thesis have to be?

Speaker Transition

Thesis for the Masters in Computational Social Science: Masters in Computational Social Science

Reid McIlroy-Young,
Masters in Computational Social Science

March 29, 2018

THE UNIVERSITY OF CHICAGO

Question

Identify computational techniques usage in science outside of explicitly computational domains

Question

Identify computational techniques usage in science outside of explicitly computational domains

Web of Science

Table	Number of Entries
publications	57136685
abstracts	26093439
publishers	50668193
keywords	78155603
references	1085738245

Selected Subsets

	Number of Publications	Number of Sources	Number of Subjects
Economics and Business	459 992	2363	6
Psychology	366 313	1017	11
Educational Sciences	168 342	1009	3
Sociology	136 894	777	9
Political Science	95 920	609	3
Other Social Sciences	83 611	571	4
Media and Communication	63 625	518	2
Law	40 829	295	2
Full Data set	1 457 418	6893	46

Table: Summary of used data, note publications can have multiple subjects, thus the final row is not a sum

Subset Explicitly Computational

	# Comp. Publications	% Explicitly Comp.	Example of Explicitly Comp. Source
Econ And Bus	60 602	13.17	Decision Support Systems
Psychology	5364	1.46	Interacting With Computers
Educational	17 988	10.69	Computers & Education
Sociology	3975	2.90	Persuasive Technology
Poli Sci	1815	1.89	Electronic Government
Other SS	2829	3.38	Adaptive Behavior
Media And Coms	24 798	38.98	Scientometrics
Law	313	0.77	Law And The Semantic Web
Full Data Set	106 680	7.32	

Table: Distribution of explicitly computational publications, note publications can have multiple subjects, thus the final row is not a sum

Subset Explicitly Computational

	# Comp. Publications	% Explicitly Comp.	Example of Explicitly Comp. Source
Econ And Bus	60 602	13.17	Decision Support Systems
Psychology	5364	1.46	Interacting With Computers
Educational	17 988	10.69	Computers & Education
Sociology	3975	2.90	Persuasive Technology
Poli Sci	1815	1.89	Electronic Government
Other SS	2829	3.38	Adaptive Behavior
Media And Coms	24 798	38.98	Scientometrics
Law	313	0.77	Law And The Semantic Web
Full Data Set	106 680	7.32	

Table: Distribution of explicitly computational publications, note publications can have multiple subjects, thus the final row is not a sum

Example of data

Field	Value
ID	WOS:000318886700008
Explicitly Computational	False
Source	APPLICATION AND BEST PRACTICE OF COMPETITIVE TECHNICAL INTELLIGENCE
Subject	Media and Communication
Year of Publications	2010
Title	Research on the Key-technology Selection of Virtual Reality Based on Patent Citation Analysis
Abstract	Based on the data of "Derwent Innovation Index" from 1963 to 2009, the authors construct a patent analysis dataset of virtual reality by retrieving the patents through keywords ...

Figure: Example of non-traditional CSS

LSTM

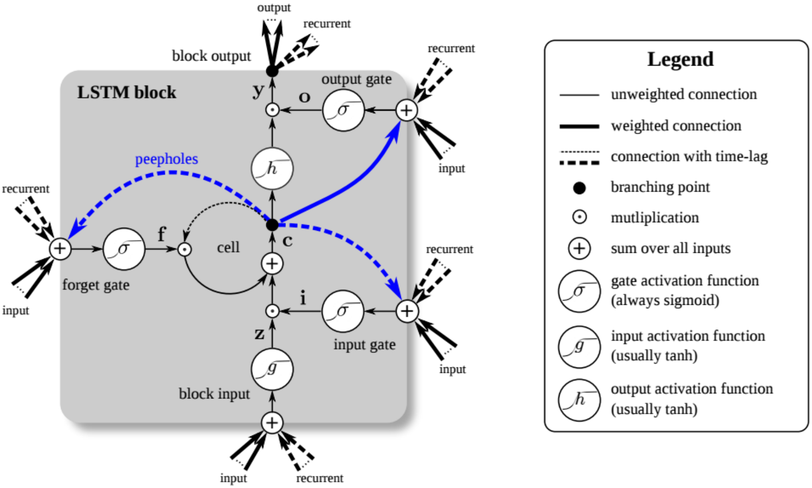


Figure: Image by Klaus Greff and colleagues as published in LSTM: A Search Space Odyssey

LSTM

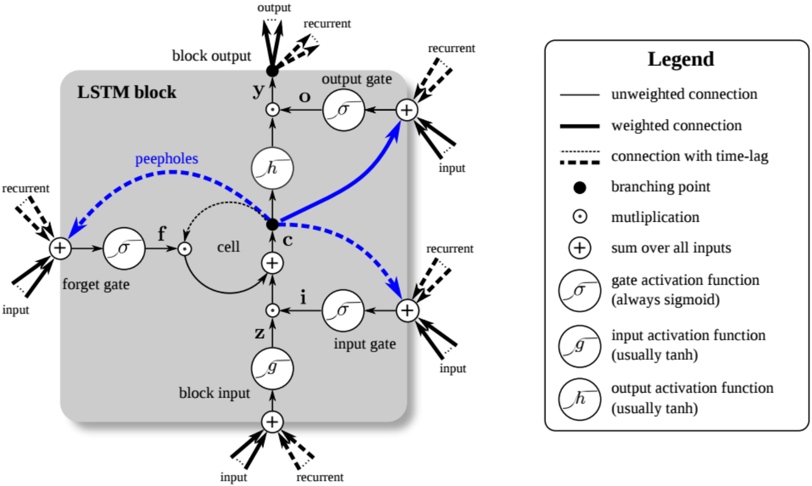
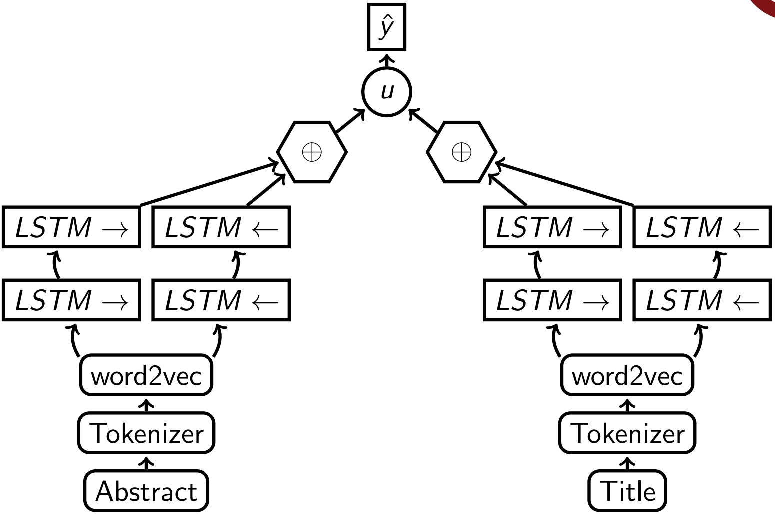
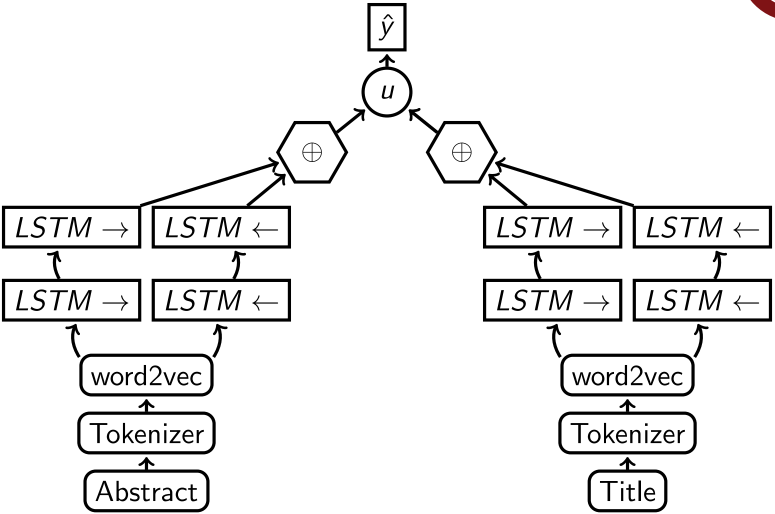


Figure: Image by Klaus Greff and colleagues as published in LSTM: A Search Space Odyssey

Simplified RNN architecture



Simplified RNN architecture



Cross Validation Results

Figure: Testing error for all subject's models, across all epochs (1 epoch is 500 training exposures), note that some data for Psychology was lost

Predicted Distributions

	% Explicitly Comp.	% Predicted Comp.	Difference
Psychology	1.50	7.90	6.40
Educational Sciences	10.70	27.50	16.80
Sociology	2.90	7.10	4.20
Political Science	1.90	12.40	10.50
Other Social Sciences	3.40	19.50	16.10
Media and Communication	39.00	36.10	−2.90
Law	0.80	5.70	4.90
Economics and Business	13.20	36.80	23.60

Table: Comparison ratios of explicitly computational publications and those predicted to be computational

Sociology Titles Wordclouds

Sociology

Computational Publications

Noncomputational Publications

Yearly Trends

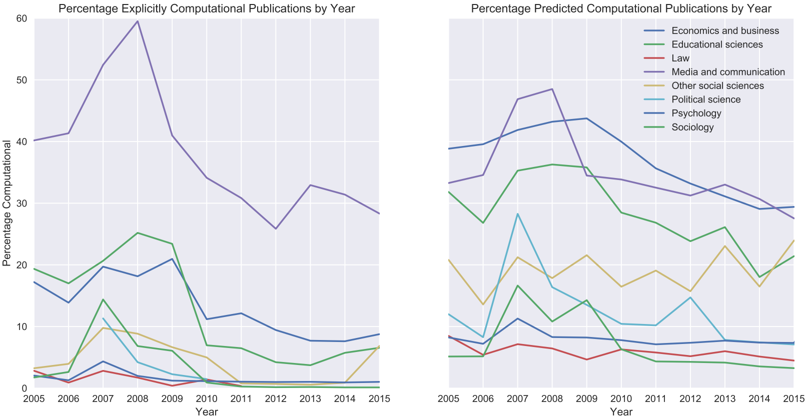


Figure: Yearly percentages of explicitly computational and predicted computational papers, notice the higher base line for the predicted papers and the smoother yearly transitions

Activations Across Words in Title

Conclusion

- Can identify CSS usage across disciplines
- There is a baseline amount of CSS usually above the explicit level
- The model is generalisable and mostly limited by computational resources

Conclusion

- Can identify CSS usage across disciplines
- There is a baseline amount of CSS usually above the explicit level
- The model is generalisable and mostly limited by computational resources

Next

- Further validation
- Change classification to ranking
- Expand scope: temporal, discipline, *etc.*
- Use attention architecture to get further insights

Speaker Transition

Proportional Improvement: An Optimal K

A Novel Metric to Uncover the Best Number of SKATER Regions

Erin M. Ochoa — emo@uchicago.edu

Spatially Constrained Clustering

A way to divide a geographic area into distinct, contiguous regions based on the characteristics of the underlying areal units

SKATER

- Spatial Kluster Analysis by Tree-Edge Removal
- A deterministic spatially constrained clustering heuristic
- Based on the minimum spanning tree
    - A path through the connectivity diagram that minimizes the total cost to traverse the graph
        - Edge cost: For two neighboring nodes, the between-node Euclidean distance in attribute space
    - Visits each node once (no circuits)
- The analyst selects the number of regions
- Cuts the minimum spanning tree into regions
    - Maximizes intra-region homogeneity & inter-region heterogeneity

SKATER

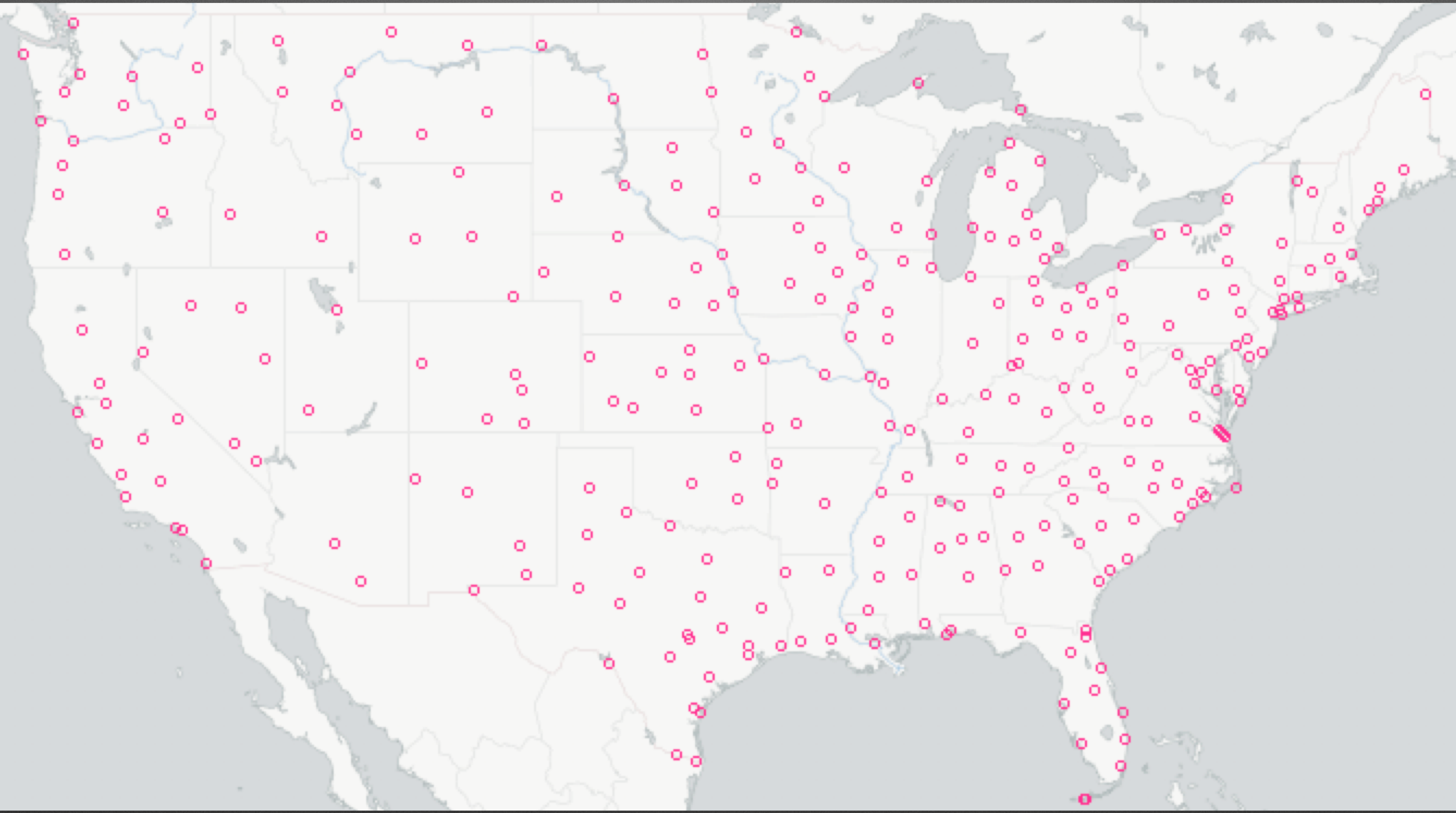
- Spatial Kluster Analysis by Tree-Edge Removal
- A deterministic spatially constrained clustering heuristic
- Based on the minimum spanning tree
  - A path through the connectivity diagram that minimizes the total cost to traverse the graph
    - Edge cost: For two neighboring nodes, the between-node Euclidean distance in attribute space
  - Visits each node once (no circuits)
- The analyst selects the number of regions
- Cuts the minimum spanning tree into regions
  - Maximizes intra-region homogeneity & inter-region heterogeneity

Research Question

When implementing spatially constrained clustering in SKATER, how can we determine the best number of regions?

Climate Regionalization

- Start with weather stations (*N* = 302)
- Convert to Thiessen polygons
- Define neighbors: First-order Queen contiguity
- Create minimum spanning tree
- Permute all possible SKATER models (*K* = [2,301])
- Compute mean squared error for each SKATER model
- Use proportional improvement to find the best *K*

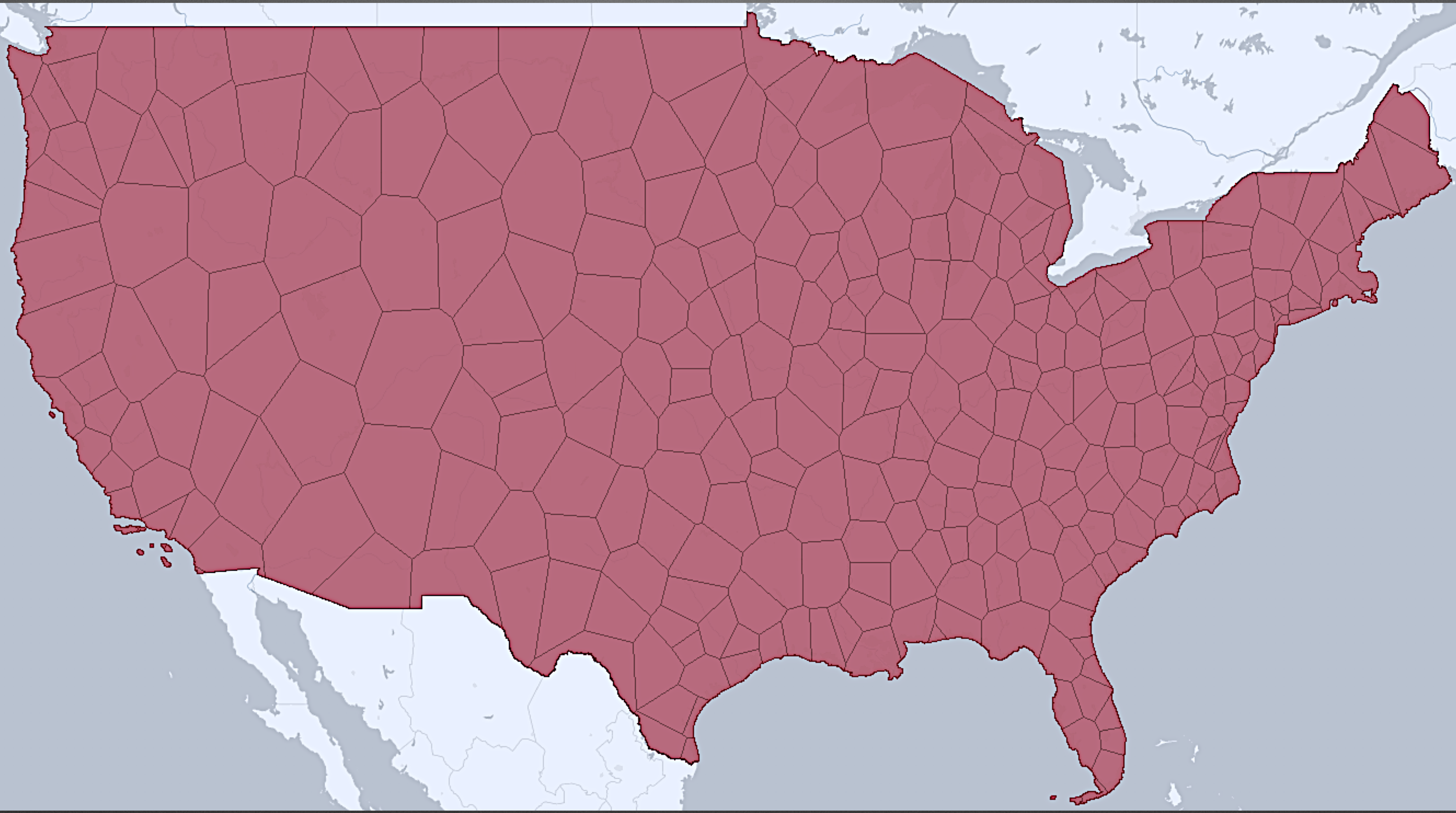


302 weather stations: 1981–2010 U.S. Climate normals from NOAA

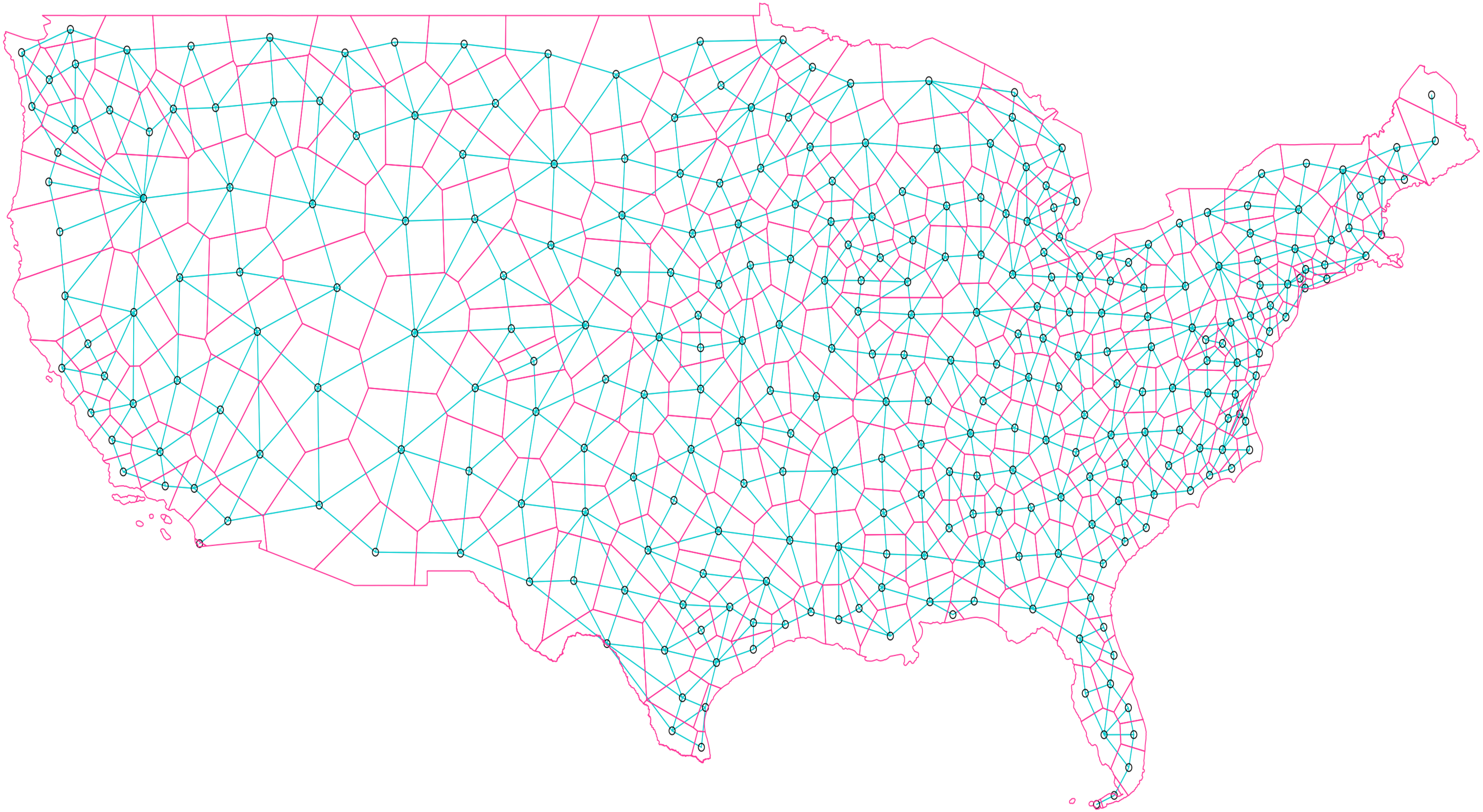
	Constant	Per-Month* Mean	Per-Month* Min. Mean	Per-Month* Max. Mean
Elevation (meters)	✓			
Daylength (minutes)		✓		
Rainfall (hundredths of inches)		✓		
Snowfall (tenths of inches)		✓		
Temperature (degrees Fahrenheit)		✓	✓	✓
Dew point (degrees Fahrenheit)		✓	✓	✓
Heat index (degrees Fahrenheit)		✓		✓
Wind chill (degrees Fahrenheit)		✓	✓	
Wind speed (miles per hour)		✓	✓	✓
Winds: % calm		✓	✓	✓
Barometric pressure (millibars)		✓	✓	✓
Clouds: % clear		✓	✓	✓
Clouds: % overcast		✓	✓	✓

Total of 337 variables for each weather station.

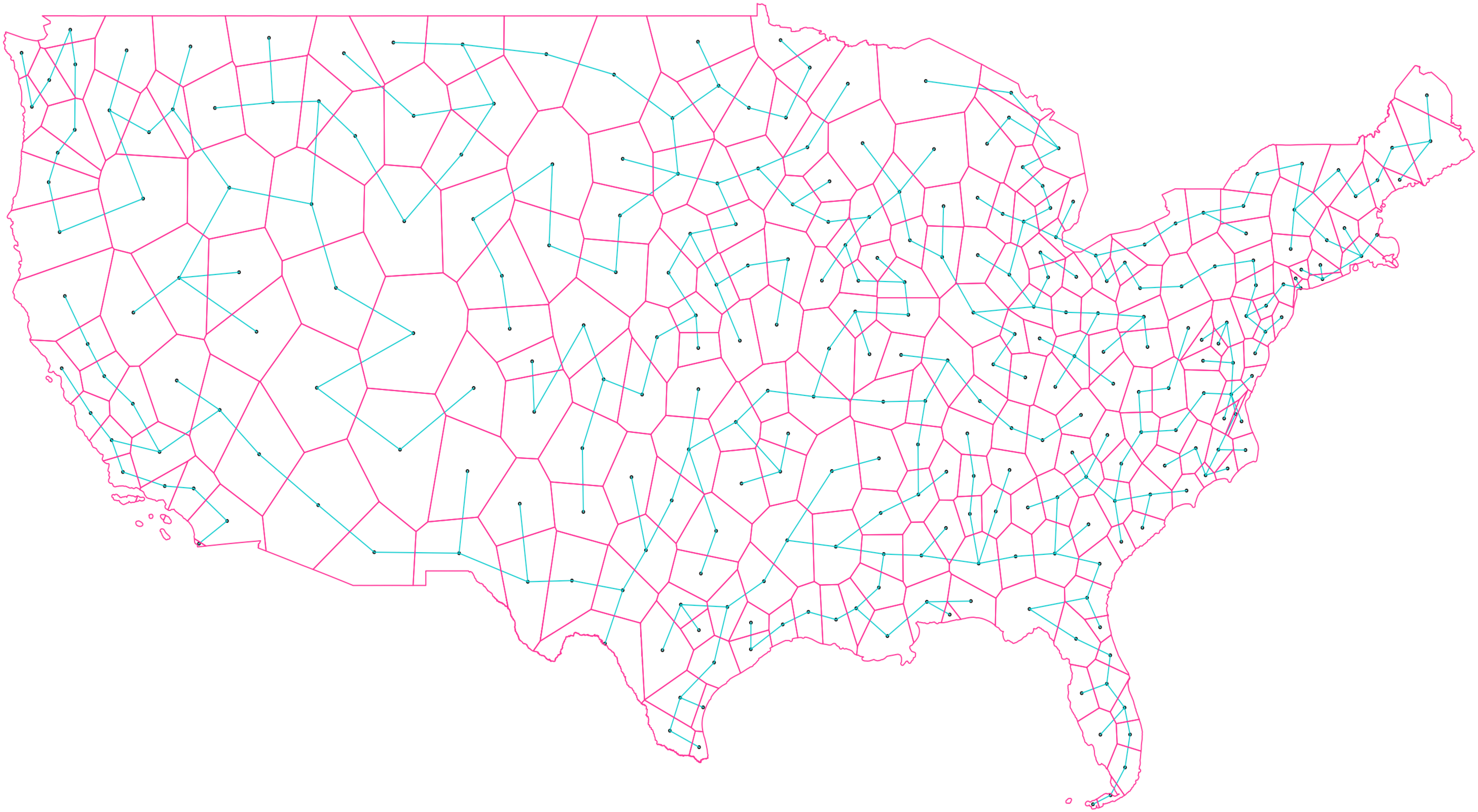
* There are twelve monthly variables, January through December.



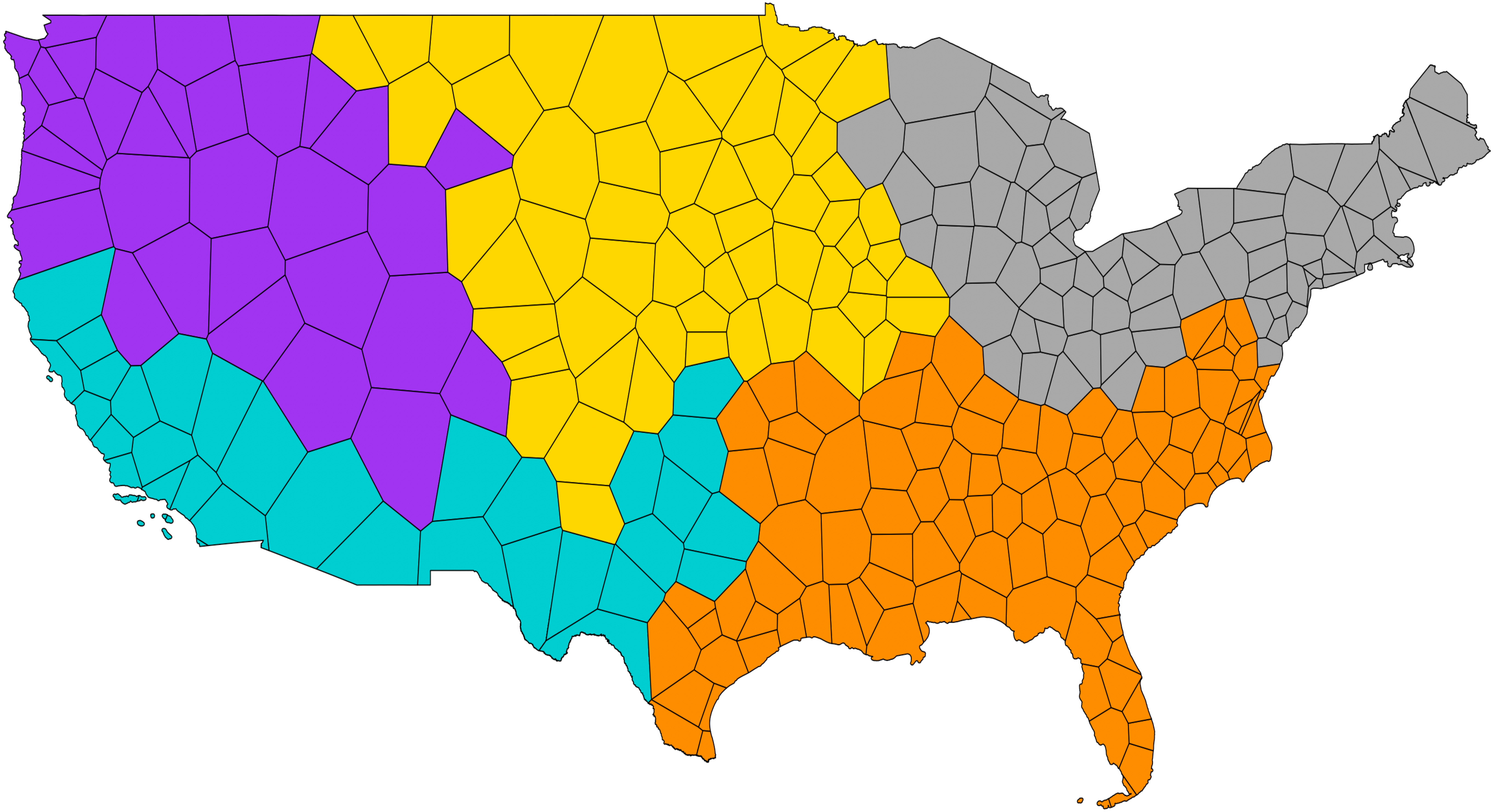
302 weather stations converted to Thiessen polygons



Connectivity diagram, Queen contiguity (first order)



Minimum spanning tree, Queen contiguity (first order)



5 climate regions as classified with SKATER

Mean Intra-Region Dissimilarity

- For a SKATER region *R* with *m* polygons (*p*), mean intra-region dissimilarity, *D*, is defined:

$$D(R) = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} d(p_i, p_j)}{m}$$

- Where $d(p_i, p_j)$ is Euclidean distance in the high-dimensional attribute space of the climate variables

Regional Error

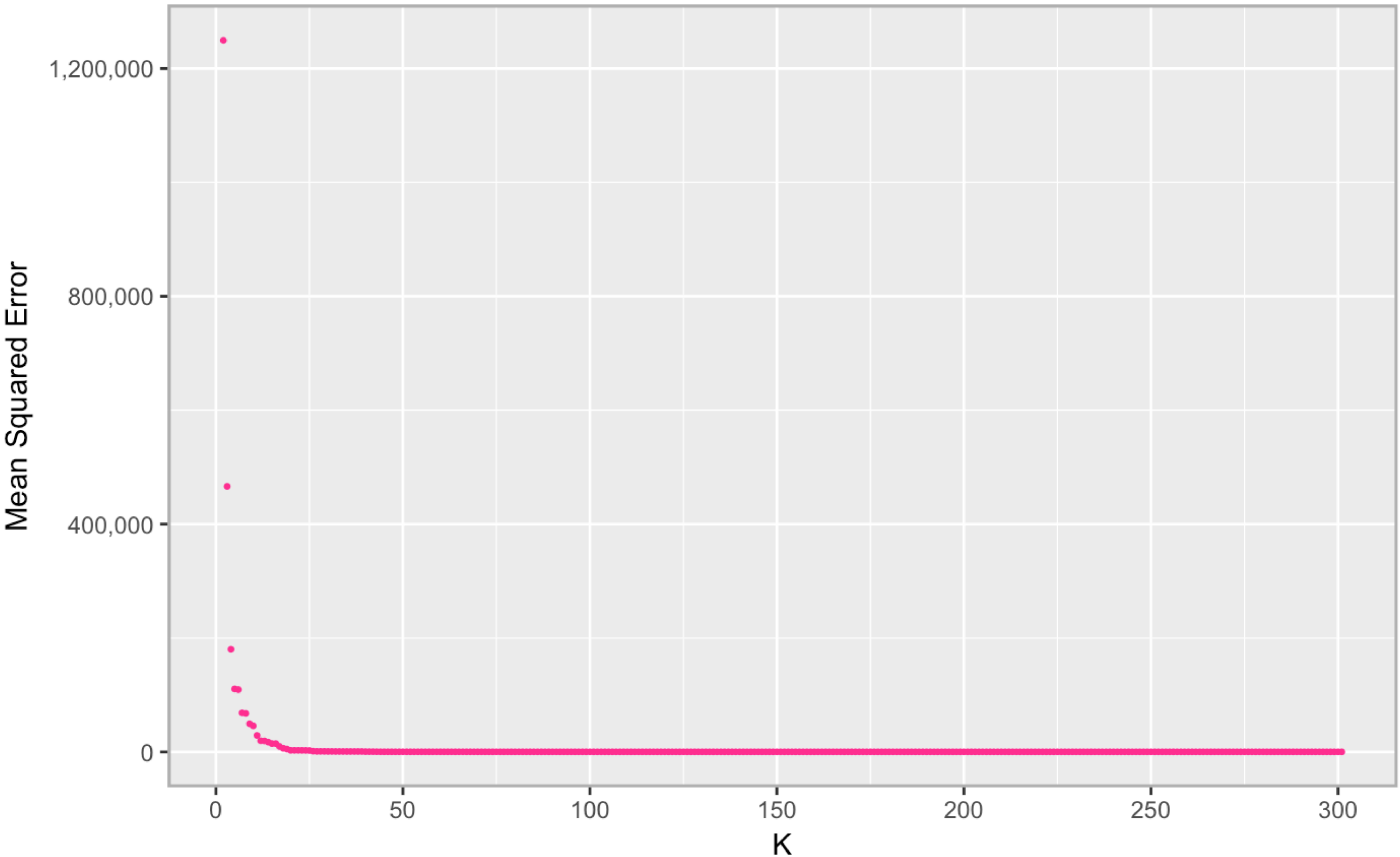
—For a region *R* with *m* polygons (*p*), the error, *E*, is defined:

$$E(R) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \Big(d(p_i, p_j) - D(R) \Big)$$

Mean Squared Error

— For a SKATER model with *K* regions, the MSE is defined:

$$MSE(K) = \frac{\sum_{R=1}^{K} E(R)^2}{K}$$

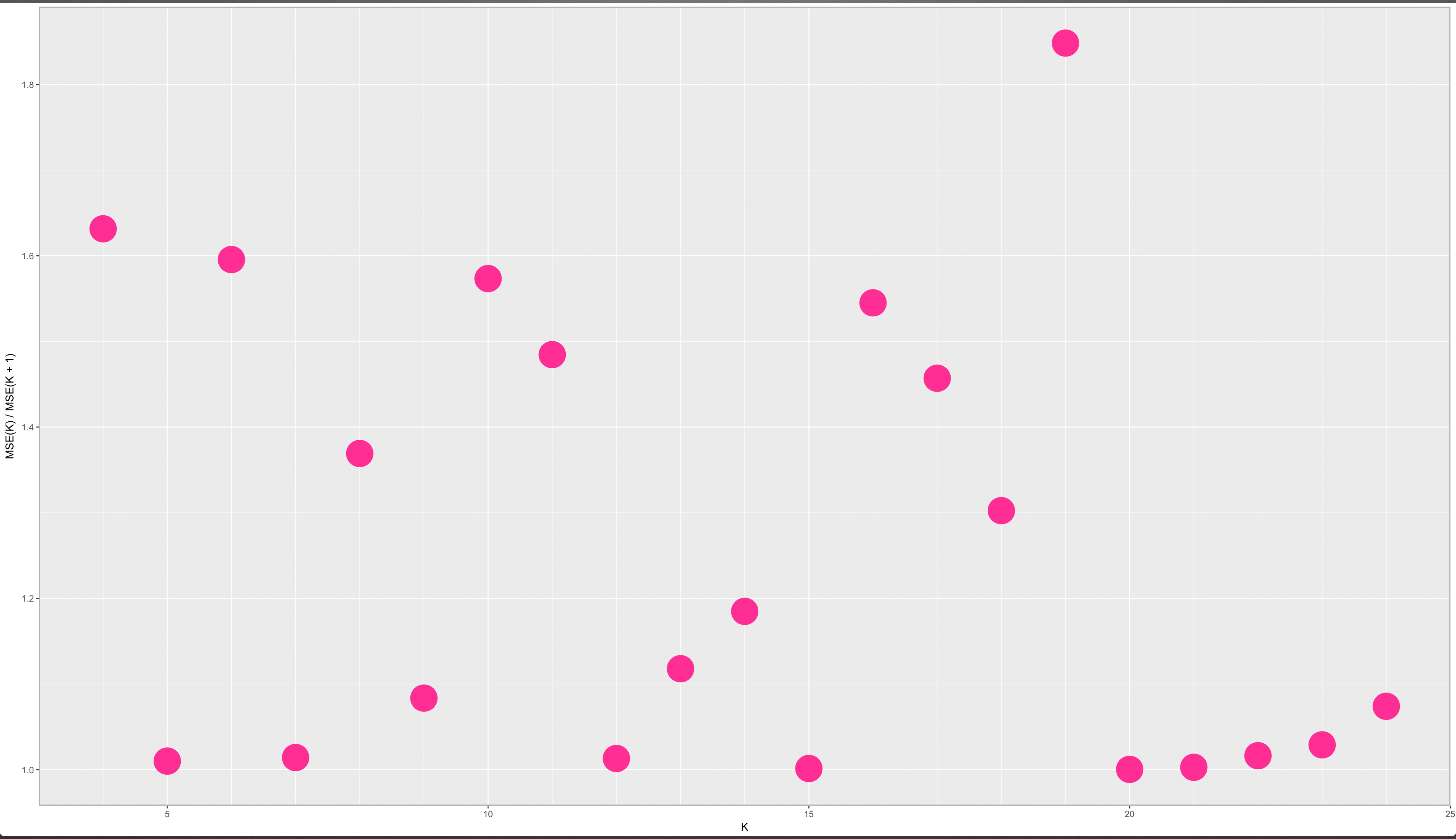


Mean squared error for SKATER models for *K* = [2,301]

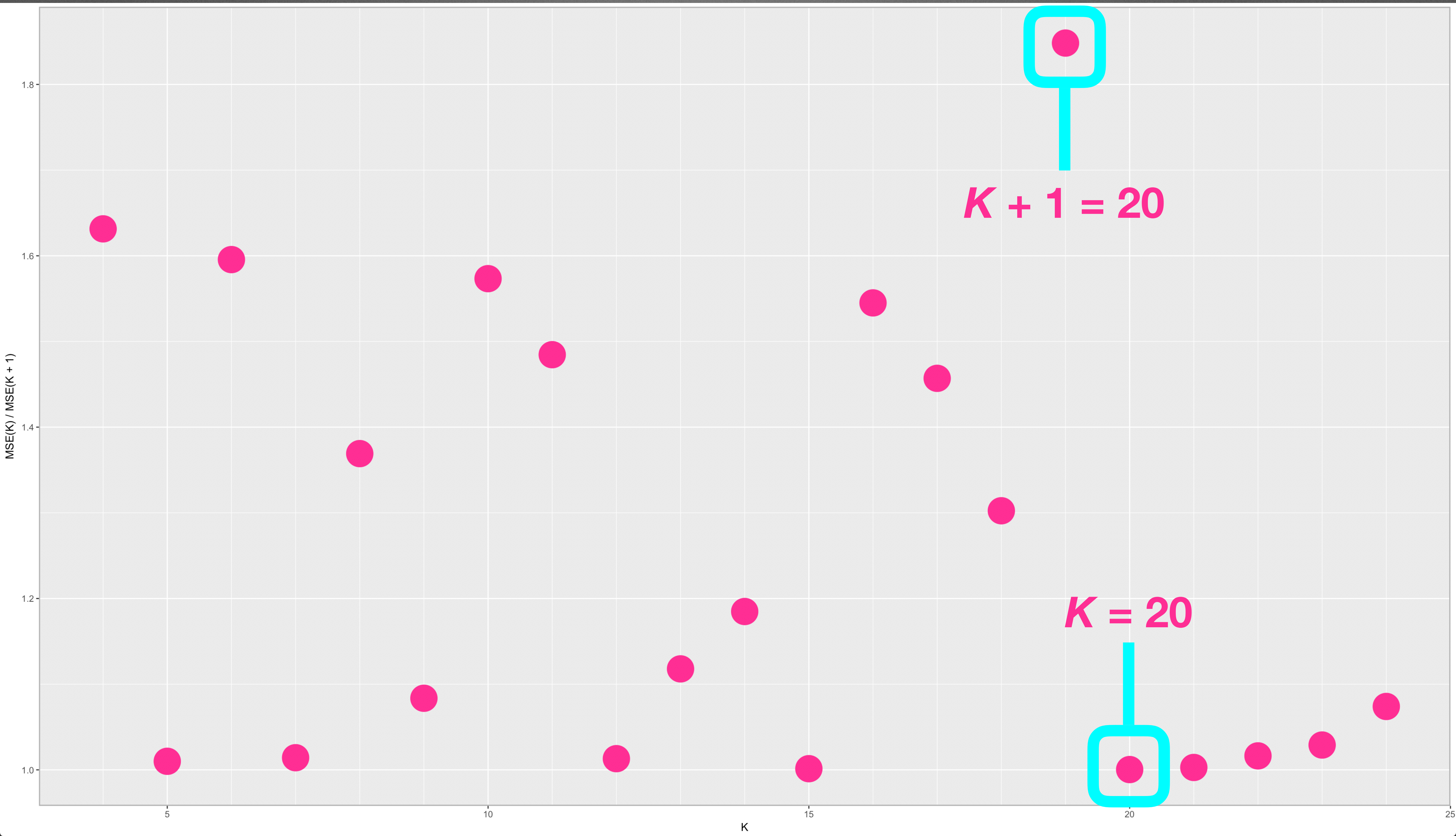
Proportional Improvement

- The proportional improvement, I, of moving from a model with K regions to a model with $K + 1$ regions is defined:

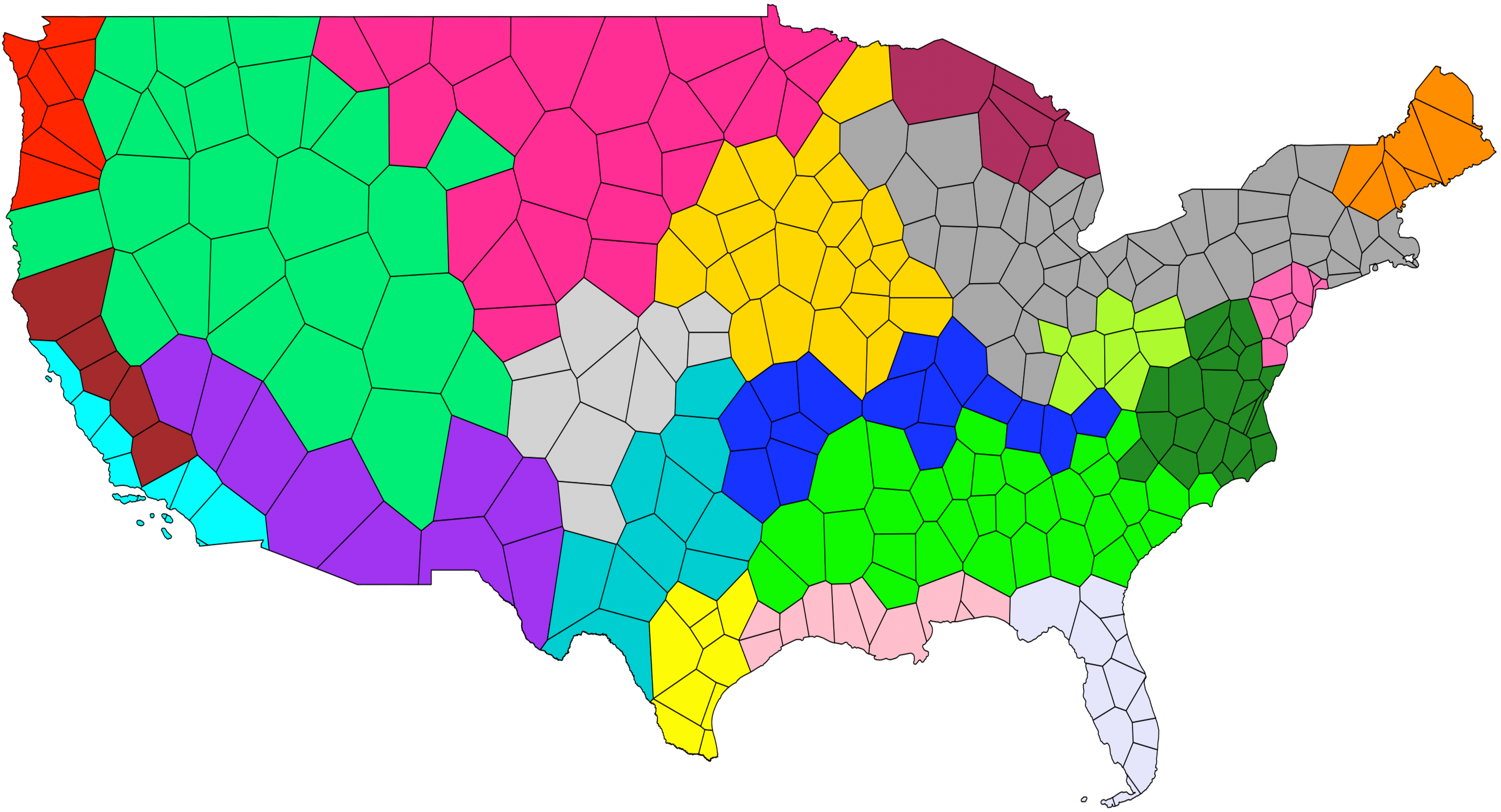
$$I(K) = \frac{MSE(K)}{MSE(K+1)}$$



Improvement in MSE gained from moving from a model with K regions to a model with $K + 1$ regions for $K = [4,24]$



Improvement in MSE gained from moving from a model with *K* regions to a model with *K* + 1 regions for *K* = [4,24]

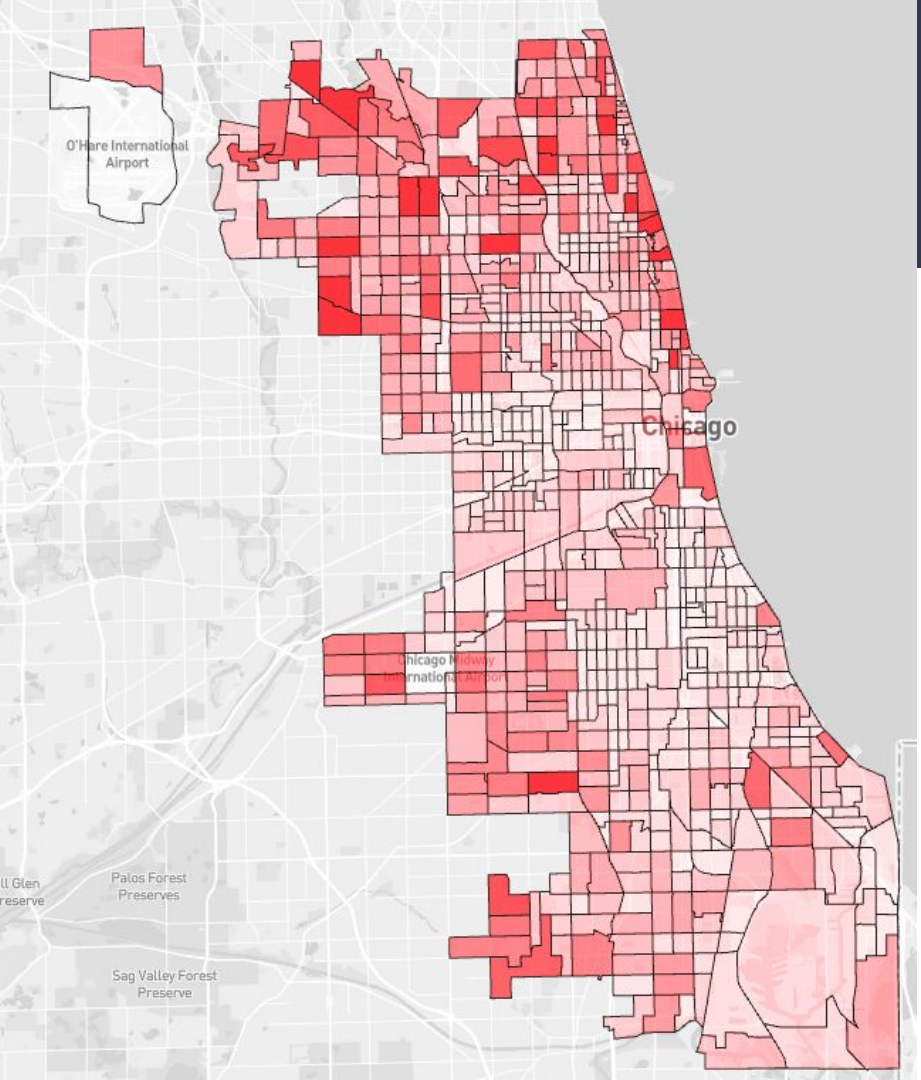


20 climate regions as classified with SKATER

Speaker Transition

An Analysis of Growth and Connection in Chicago's Neighborhoods

By: Benjamin Rothschild
Advised by: Luis Bettencourt & Daniel Zuend



Research Question

How are the neighborhoods of Chicago structured and how do people and money flow between them?

How do these connections between places influence the growth of the city?

Longitudinal Employer-Household Dynamics (LEHD)

- Synthetic dataset
- States supply data on employment, wages
- Census provides data on where people live and demographic info
- These datasets are joined and the to create census block-level data on where people live and work.

Longitudinal Employer–Household Dynamics (LEHD)

- Synthetic dataset
- States supply data on employment, wages
- Census provides data on where people live and demographic info
- These datasets are joined and the to create census block-level data on where people live and work.

Dataset

	w_geocode	h_geocode	S000	SA01	SA02	SA03	SE01	SE02	SE03	SI01	SI02	SI03
0	170010001001001	170010104003020	1	0	1	0	0	0	1	0	0	1
1	170010001001003	170010011001025	1	0	1	0	0	0	1	0	1	0
2	170010001001003	170010011002036	1	0	1	0	0	0	1	0	1	0
3	170010001001003	170010101004121	1	1	0	0	0	1	0	0	1	0
4	170010001001003	170010106003018	1	0	0	1	0	0	1	0	1	0

S000 = Total Jobs
SA01 = Workers age 29 younger
SA02 = Workers age 30 - 54
SA03 = Workers over 55

SE01 = Earning $1,250/month less
SE02 = Earning $1,250 - $3,333
SE03 = Earning over $3,333
SI01 = Goods producing
SI02 = Trade Producing
SI03 = Other

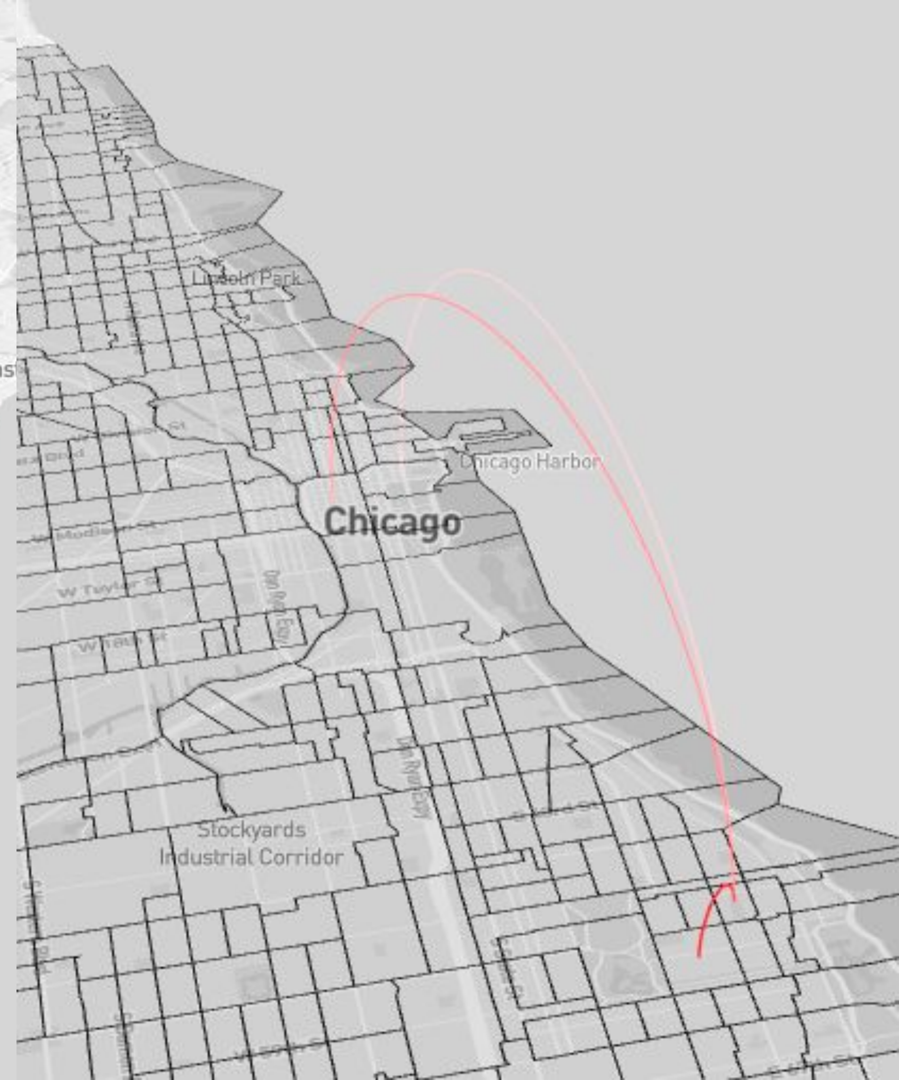
Dataset

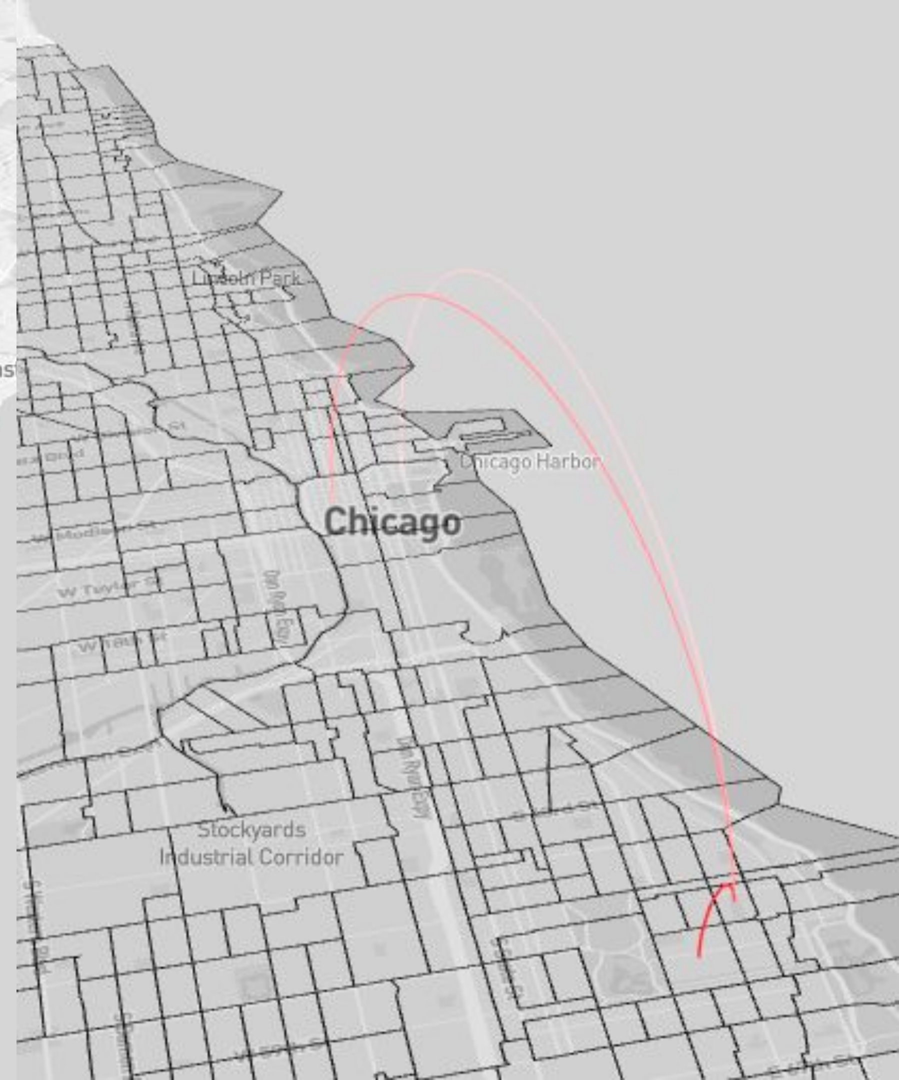
	w_geocode	h_geocode	S000	SA01	SA02	SA03	SE01	SE02	SE03	SI01	SI02	SI03
0	170010001001001	170010104003020	1	0	1	0	0	0	1	0	0	1
1	170010001001003	170010011001025	1	0	1	0	0	0	1	0	1	0
2	170010001001003	170010011002036	1	0	1	0	0	0	1	0	1	0
3	170010001001003	170010101004121	1	1	0	0	0	1	0	0	1	0
4	170010001001003	170010106003018	1	0	0	1	0	0	1	0	1	0

S000 = Total Jobs
SA01 = Workers age 29 younger
SA02 = Workers age 30 - 54
SA03 = Workers over 55

SE01 = Earning $1,250/month less
SE02 = Earning $1,250 - $3,333
SE03 = Earning over $3,333
SI01 = Goods producing
SI02 = Trade Producing
SI03 = Other


Elgin
DuPage County Airport
Busse Woods
Chicago Executive Airport
O'Hare International Airport
Chicago Midway International Airport
Chicago


Lincoln Park
Chicago Harbor
Chicago
Stockyards Industrial Corridor


Elgin
DuPage County Airport
Busse Woods
Chicago Executive Airport
O'Hare International Airport
Evans
Chicago Midway International Airport
Chicago

Lincoln Park
Chicago Harbor
Chicago
Stockyards Industrial Corridor

Convert Dataset to Matrices

h_tract w_tract	17031010100	17031010201	17031010202	17031010300	17031010400
17031010100	0.041971	0.011354	0.009747	0.010343	0.003482
17031010201	0.004562	0.010543	0.017544	0.003232	0.001393
17031010202	0.006387	0.017843	0.017544	0.006464	0.000696
17031010300	0.018248	0.008921	0.003899	0.034906	0.004875
17031010400	0.009124	0.005677	0.001949	0.018100	0.136490

x

h_tract	
17031010100	46.0
17031010201	14.0
17031010202	5.0
17031010300	16.0
17031010400	5.0

=

w_tract	S000
17031010100	1
17031010100	1
17031010100	1
17031010100	1
17031010100	1

Convert Dataset to Matrices

h_tract w_tract	17031010100	17031010201	17031010202	17031010300	17031010400
17031010100	0.041971	0.011354	0.009747	0.010343	0.003482
17031010201	0.004562	0.010543	0.017544	0.003232	0.001393
17031010202	0.006387	0.017843	0.017544	0.006464	0.000696
17031010300	0.018248	0.008921	0.003899	0.034906	0.004875
17031010400	0.009124	0.005677	0.001949	0.018100	0.136490

x

h_tract	
17031010100	46.0
17031010201	14.0
17031010202	5.0
17031010300	16.0
17031010400	5.0

=

w_tract	S000
17031010100	1
17031010100	1
17031010100	1
17031010100	1
17031010100	1

Calculate Most Influential Tracts using Eigenvectors

Eigenvector Centrality

We assign relative scores to all tracts based on it’s connections.

Connections to influential tracts will give you a higher score than a connections to lower tracts

This concept is similar to the PageRank algorithm which was the basis of Google’s search algorithm which ranked web pages based on hyperlinks

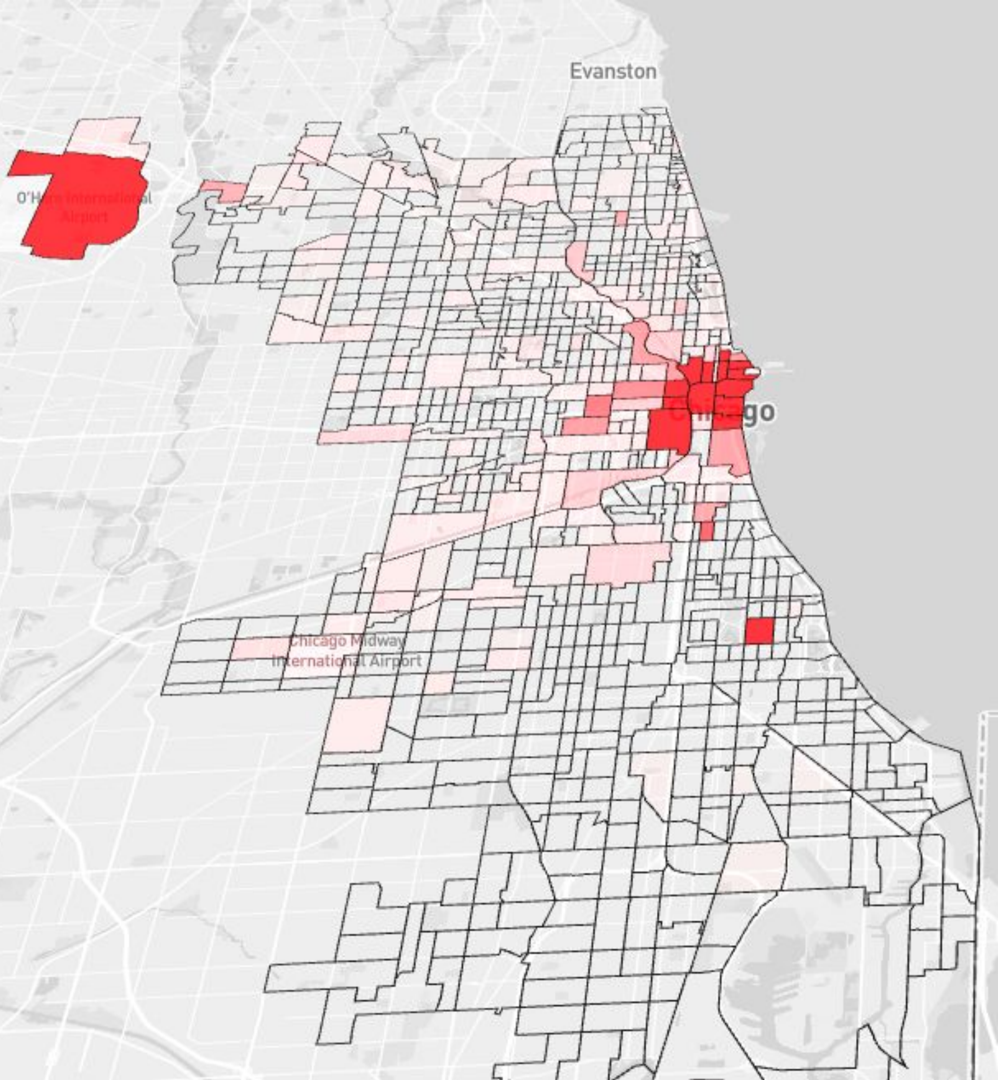
Calculate Most Influential Tracts using Eigenvectors

Eigenvector Centrality

We assign relative scores to all tracts based on it’s connections.

Connections to influential tracts will give you a higher score than a connections to lower tracts

This concept is similar to the PageRank algorithm which was the basis of Google’s search algorithm which ranked web pages based on hyperlinks

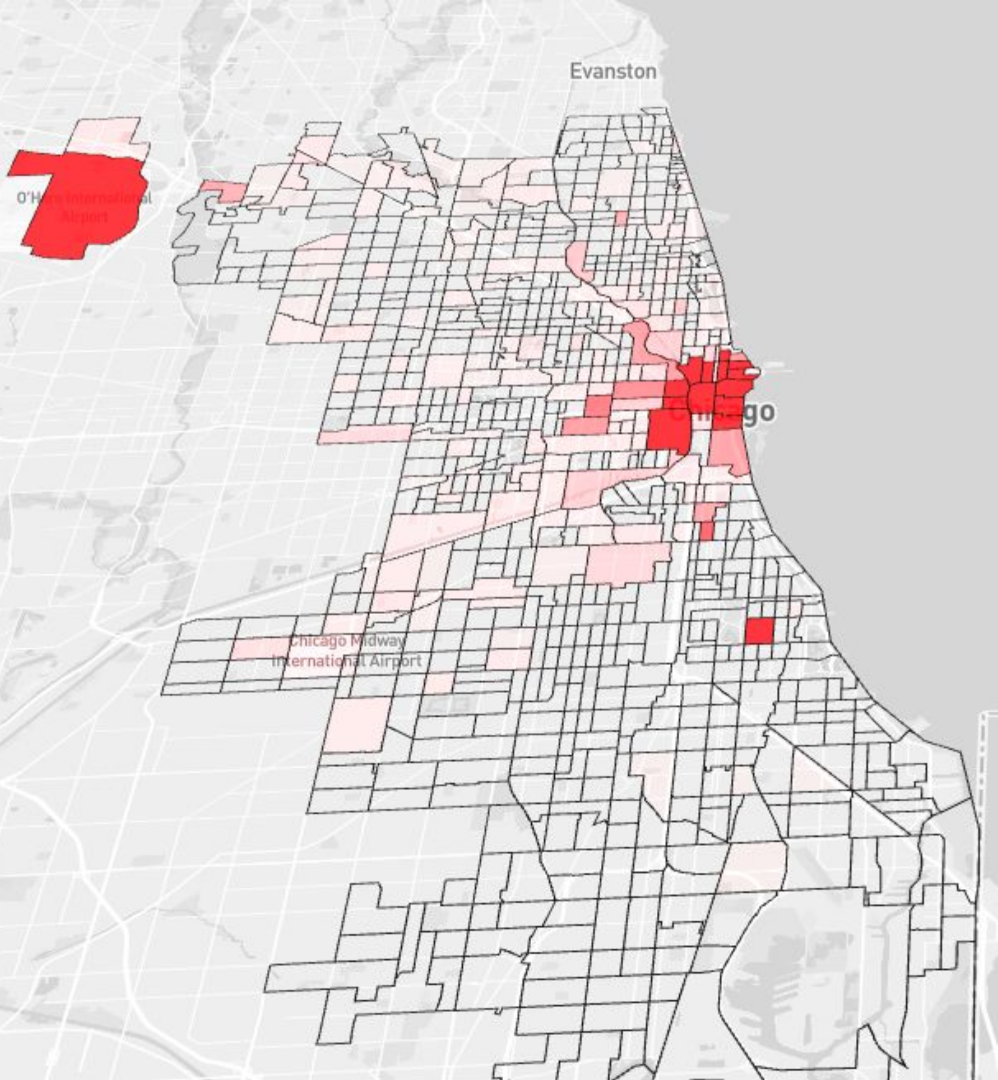


"Gould Index of Accessibility"

Relative number of paths joining each vertex to all vertices in the graph

Equilibrium relative importance of vertices

The equilibrium distribution of a rumor spreading in the graph from any vertex

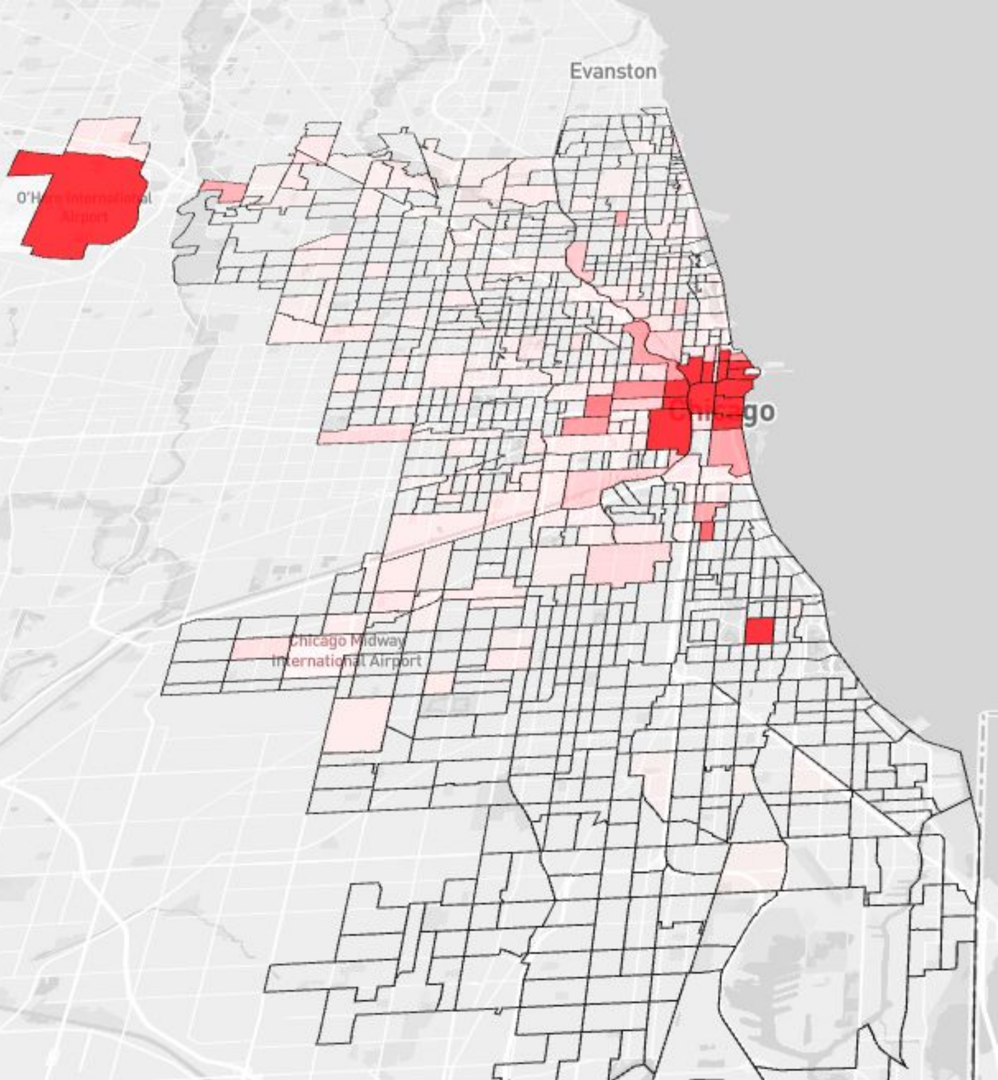


"Gould Index of Accessibility"

Relative number of paths joining each vertex to all vertices in the graph

Equilibrium relative importance of vertices

The equilibrium distribution of a rumor spreading in the graph from any vertex

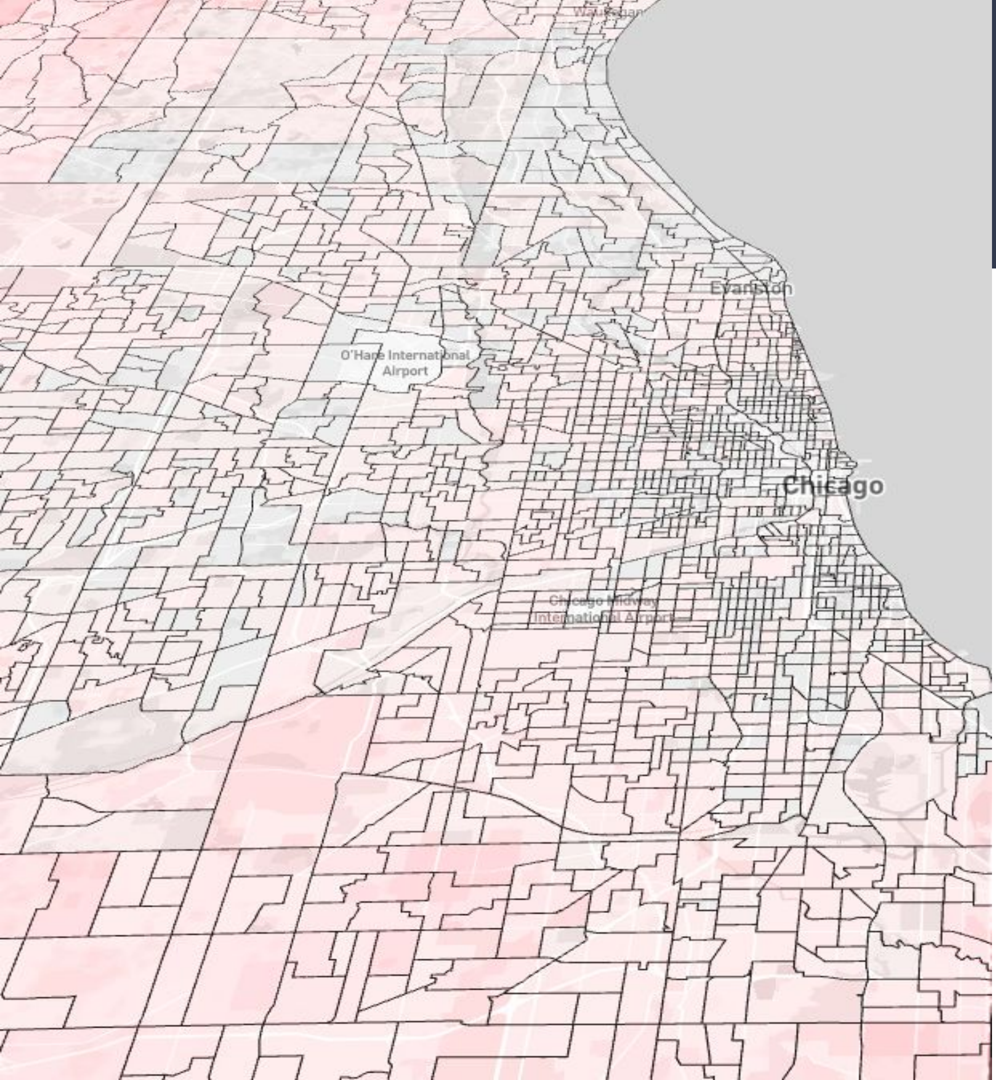


"Gould Index of Accessibility"

Relative number of paths joining each vertex to all vertices in the graph

Equilibrium relative importance of vertices

The equilibrium distribution of a rumor spreading in the graph from any vertex



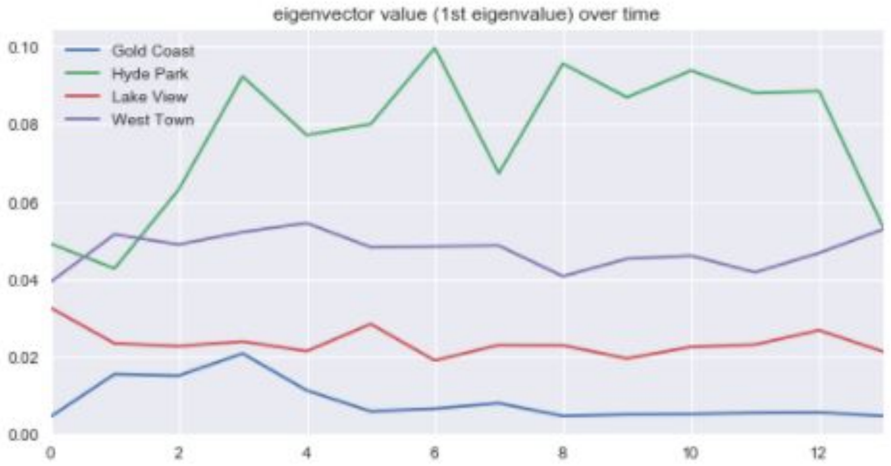
"Gould Index of Accessibility"

Relative number of paths joining each vertex to all vertices in the graph

Equilibrium relative importance of vertices

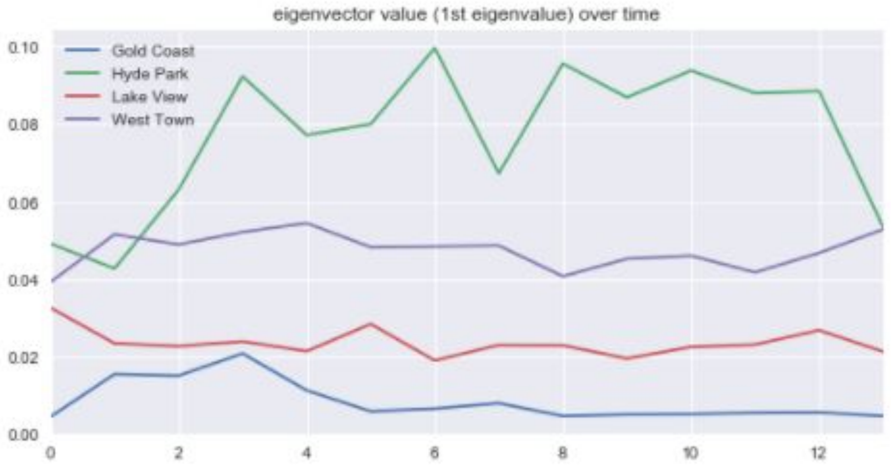
The equilibrium distribution of a rumor spreading in the graph from any vertex

Further Discussions



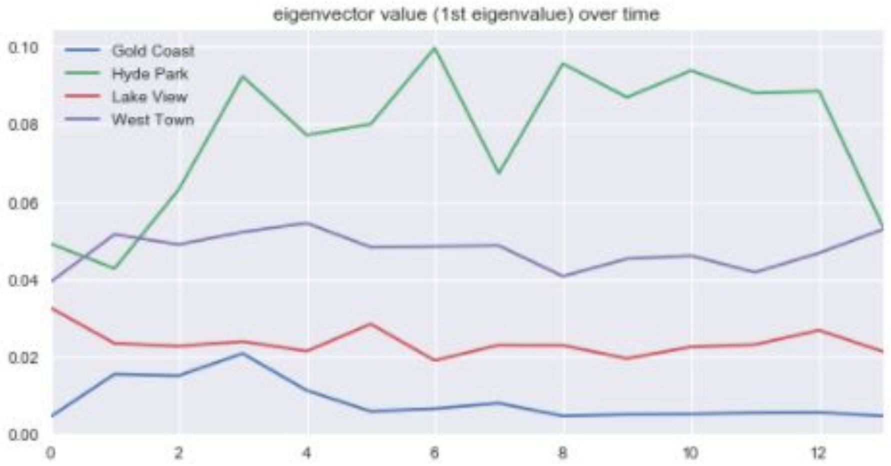
1. How does the importance of neighborhoods in the network change over time?
2. How predictive are these network accessibility indices to neighborhood investments building permits and house values?
3. How do will investments in one neighborhood affect another?

Further Discussions



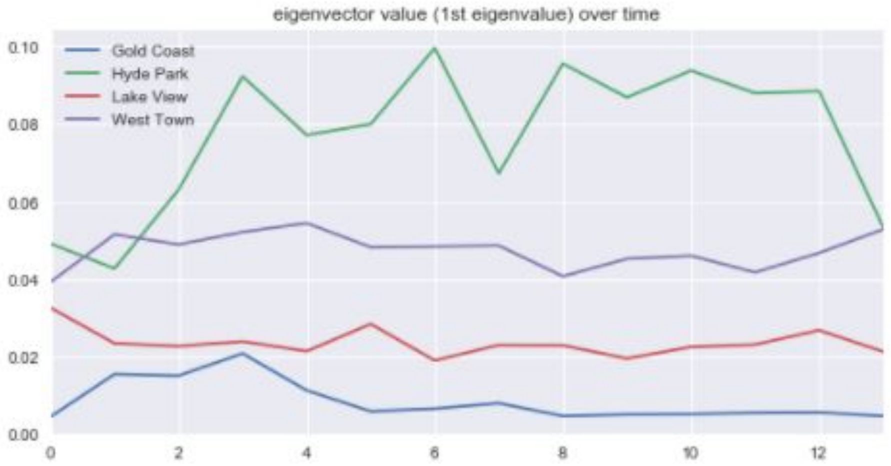
1. How does the importance of neighborhoods in the network change over time?
2. How predictive are these network accessibility indices to neighborhood investments building permits and house values?
3. How do will investments in one neighborhood affect another?

Further Discussions



1. How does the importance of neighborhoods in the network change over time?
2. How predictive are these network accessibility indices to neighborhood investments building permits and house values?
3. How do will investments in one neighborhood affect another?

Further Discussions



1. How does the importance of neighborhoods in the network change over time?
2. How predictive are these network accessibility indices to neighborhood investments building permits and house values?
3. How do will investments in one neighborhood affect another?

Speaker Transition

The Impact of Postponing Announced Unconventional Fiscal Policy on Consumption Expenditure

Xinzhu Sun

March 2018

THE UNIVERSITY OF CHICAGO

Introduction

- The basic issue I address is the adjustment of reaction and expectation of Japanese consumers when government announces a future sales tax increase but has the option to revise it.

Introduction

- The basic issue I address is the adjustment of reaction and expectation of Japanese consumers when government announces a future sales tax increase but has the option to revise it.
- First, I have my theoretical model built on New Keynesian theory. Then I use Japanese quarterly Macroeconomics data to train my model and do simulation.

What is Unconventional Fiscal Policy

- Eggertsson and Woodford (2006) recently proposed unconventional fiscal policy as an alternative to stimulating demand by changing inter-temporal prices.

What is Unconventional Fiscal Policy

- Eggertsson and Woodford (2006) recently proposed unconventional fiscal policy as an alternative to stimulating demand by changing inter-temporal prices.
- "An unconventional fiscal policy uses announcements of future increases in consumption taxes to generate inflation expectations and accelerate consumption expenditure."

What is Unconventional Fiscal Policy

- Eggertsson and Woodford (2006) recently proposed unconventional fiscal policy as an alternative to stimulating demand by changing inter-temporal prices.
- "An unconventional fiscal policy uses announcements of future increases in consumption taxes to generate inflation expectations and accelerate consumption expenditure."

Value-added Tax Policy sequence in Japan

-

Value-added Tax Policy sequence in Japan

- In June 2012, the Noda government announced to raise the consumption tax to 8% in April 2014 and to 10% in October 2015.

Value-added Tax Policy sequence in Japan

- In June 2012, the Noda government announced to raise the consumption tax to 8% in April 2014 and to 10% in October 2015.
- In November 2014, the Abe government announced to postpone the second tax increase to April 2017.

Value-added Tax Policy sequence in Japan

- In June 2012, the Noda government announced to raise the consumption tax to 8% in April 2014 and to 10% in October 2015.
- In November 2014, the Abe government announced to postpone the second tax increase to April 2017.
- In May 2016, a second postponement was announced by Abe government, which further delayed the tax increase to October 2019.

Research Design

- Use a standard new Keynesian model similar to the one analyzed by Farhi et al.(2013) while introducing the idea of time-inconsistency into the model and introduce the autoregressive conditional density process (ARCD) into my model.

The reduced system of a standard new Keynesian model

$$\tilde{y}_t = E_t \tilde{y}_{t+1} - \tilde{r}_{t+1}$$

$$\tilde{\pi}_t = \frac{(1-\phi)(1-\phi\beta)}{\phi} \kappa (\tilde{y}_t - \tilde{y}_t^f) + \beta E_t \tilde{\pi}_{t+1}$$

$$\tilde{y}_t^f = \rho \tilde{a}_{t-1}^f + \epsilon_{a,t}$$

$$-\nu \tilde{m}_t = -\tilde{y}_t + (\frac{1}{i * (1 + i*)})(\tilde{r}_{t+1} + E_t \tilde{\pi}_{t+1})$$

$$\Delta \tilde{m}_t = (1-\rho_m)\phi * - \tilde{\pi}_t + \rho_m \tilde{\pi}_{t-1} + \rho_m \Delta \tilde{m}_{t-1} + \epsilon_{m,t}$$

$$\Delta \tilde{m}_t = \tilde{m}_t - \tilde{m}_{t-1}$$

The expected tax rate updates following an AECD process:

$$\tau_t^C = \rho \tau_{t-1}^C + \epsilon_t$$

The reduced system of a standard new Keynesian model

$$\tilde{y}_t = E_t \tilde{y}_{t+1} - \tilde{r}_{t+1}$$

$$\tilde{\pi}_t = \frac{(1-\phi)(1-\phi\beta)}{\phi}\kappa(\tilde{y}_t - \tilde{y}_t^f) + \beta E_t \tilde{\pi}_{t+1}$$

$$\tilde{y}_t^f = \rho \tilde{a}_{t-1}^f + \epsilon_{a,t}$$

$$-\nu \tilde{m}_t = -\tilde{y}_t + (\frac{1}{i * (1+i*)})(\tilde{r}_{t+1} + E_t \tilde{\pi}_{t+1})$$

$$\Delta \tilde{m}_t = (1-\rho_m)\phi * - \tilde{\pi}_t + \rho_m \tilde{\pi}_{t-1} + \rho_m \Delta \tilde{m}_{t-1} + \epsilon_{m,t}$$

$$\Delta \tilde{m}_t = \tilde{m}_t - \tilde{m}_{t-1}$$

The expected tax rate updates following an AECD process:

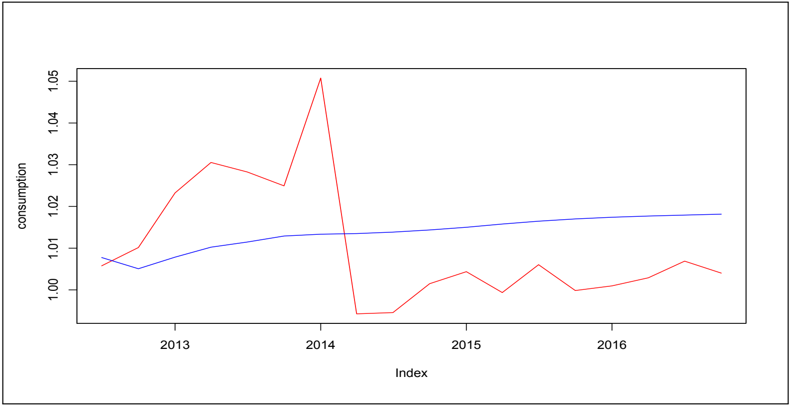
$$\tau_t^C = \rho \tau_{t-1}^C + \epsilon_t$$

Preliminary Suppositions and Implications

- First, the pre-announcement of tax increase generate high consumption in the short time, but after the first exercise day, consumption drops sharply and does not maintain at the same level if the policy were not conducted.

Preliminary Suppositions and Implications

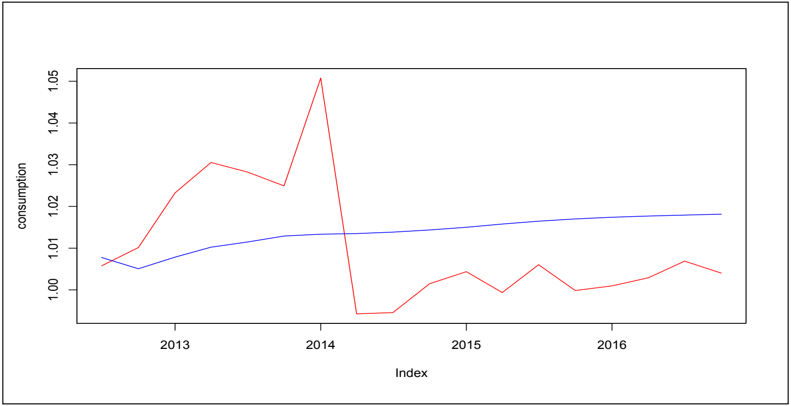
Figure: Real Consumption Index vs Predicted Consumption Indicies



Real Wold Value: Red; Predicted Value of VAR: Blue; Predicted Value of VARX: Green

Preliminary Suppositions and Implications

Figure: Real Consumption Index vs Predicted Consumption Index



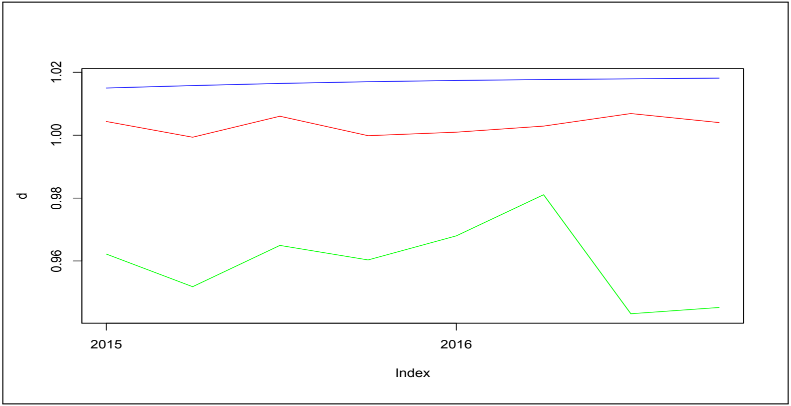
Real Wold Value: Red; Predicted Value: Blue

Preliminary Suppositions and Implications

- Second,the postponement of the second unconventional fiscal policy actually drags consumption to a higher level than if the policy were not postponed. If no postponement occurred, the unconventional policy sequence would drive the consumption expenditure to a level even worse.

Preliminary Suppositions and Implications

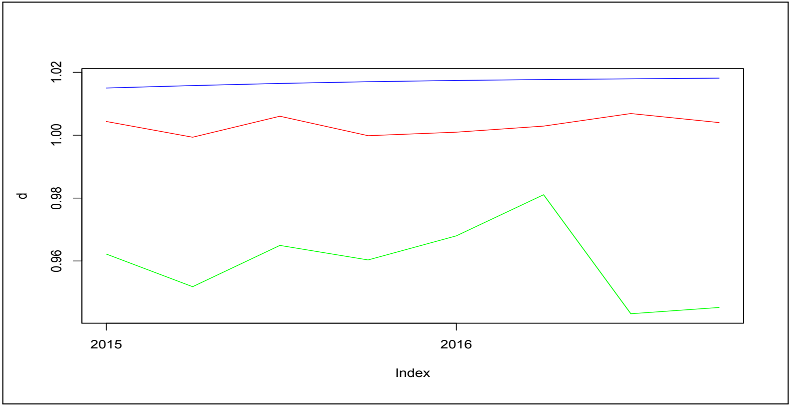
Figure: Real Consumption Index vs Predicted Consumption Indicies



Real Wold Value: Red; Predicted Value: Blue; Predicted Value of VARX: Green

Preliminary Suppositions and Implications

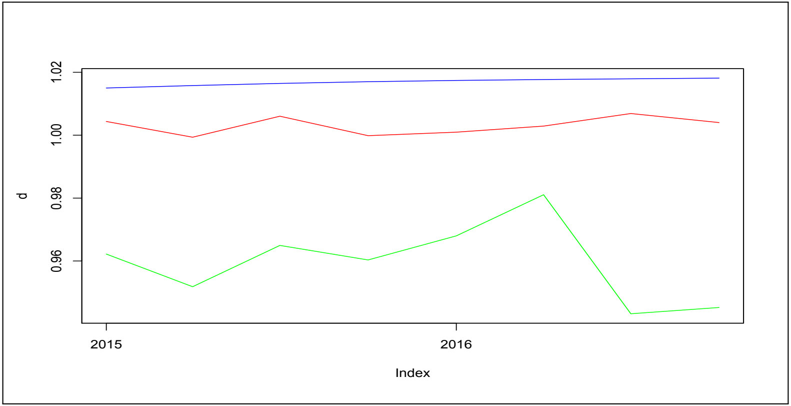
Figure: Real Consumption Index vs Predicted Consumption Indicies



Real Wold Value: Red; Predicted Value: Blue;
Predicted Value of VARX: Green

Preliminary Suppositions and Implications

Figure: Real Consumption Index vs Predicted Consumption Indicies



Real Wold Value: Red; Predicted Value: Blue;
Predicted Value of VARX: Green

Speaker Transition

Computational Social Sciences Workshop: Student Conference

Participants of Music Festivals:

Relationship between Styles of Amenities and Types of People

Xingyun Wu

3/29/2018

Background: Theory

- Amenities could drive urban development by making some cities more attractive to human capital than others (Clark, 2003)
- Different urban amenities could attract or repel different types of residents (Clark, 2003)
- Types of amenities (Clark, 2003)
  - Natural amenities, **constructed amenities**, socio-economic composition and diversity, values and attitudes of residents
- Scenes (Silver & Clark, 2016)
  - A conceptual tool to identify the range and configurations of aesthetic meanings
  - 3 components and 15 dimensions

Background: Empirical Studies

- Effect of Bohemia and art on behavior (Jeong, 2016)
  - The use of transportation data to understand relationship between amenities and human behavior

- Analyses using transportation data
  - Spatial-temporal patterns of bicycle rides (Faghih-Imani & Eluru, 2015, 2016a, 2016b; Zhou, 2015)
  - The effect of weather and calendar events on "spatial-temporal" patterns of bicycle sharing system (Corcoran et al., 2014)
  - Spatial-temporal patterns of taxi trips (Ferreira, 2013; Zhan et al., 2013; Liu, 2012; Qi et al., 2011)

Research Question

- **Main Question**
  - What styles of amenities attract what types of people?
- **Problems in Answering This Question**
  - Lack of available survey data
  - Difficulties to understand meanings/styles of amenities
  - Difficulties to link amenities and people

Research Question

- **My Solutions**
  - Special events — public cultural events (e.g. music festivals)
  - Spatial analysis with transportation data to capture patterns
  - Computational content analysis to understand meanings
- **Specific Questions**
  - What are the styles of some specific cultural events?
  - Where do the attendants of these cultural events come from?
  - Relationship between people and the cultural events?
  - Relationship between scenes?

Research Question

- **My Solutions**
  - Special events — public cultural events (e.g. music festivals)
  - Spatial analysis with transportation data to capture patterns
  - Computational content analysis to understand meanings
- **Specific Questions**
  - What are the styles of some specific cultural events?
  - Where do the attendants of these cultural events come from?
  - Relationship between people and the cultural events?
  - Relationship between scenes?

Research Design

- **Step 1: computational content analysis using description of different music types**
  - Understand meanings/styles of cultural events
- **Step 2: spatial analysis using transportation data**
  - Detect changes in transportation records made by cultural events
- **Step 3: statistical analysis using transportation data**
  - Understand relationship between basic demographic variables (age and gender) and cultural events
- **Step 4: statistical analysis using community/scenes data**
  - Understand relationship between scenes

Research Design

- **Step 1: computational content analysis using description of different music types**
  - Understand meanings/styles of cultural events
- **Step 2: spatial analysis using transportation data**
  - Detect changes in transportation records made by cultural events
- **Step 3: statistical analysis using transportation data**
  - Understand relationship between basic demographic variables (age and gender) and cultural events
- **Step 4: statistical analysis using community/scenes data**
  - Understand relationship between scenes

Data of First Three Steps

- **Webpages of Wikipedia (n = 1312)**
  - Type: textual data
  - Main Content:
    - Description of different music types
- **Divvy Bicycle System Data (n = 3,829,014, Year 2017)**
  - Type: numeric data
  - Main Variables:
    - Time: trip start time, trip end time
    - Location: latitudes and longitudes of trip origins and destinations
    - People: gender, age

Data of First Three Steps

- **Webpages of Wikipedia (n = 1312)**
  - Type: textual data
  - Main Content:
    - Description of different music types
- **Divvy Bicycle System Data (n = 3,829,014, Year 2017)**
  - Type: numeric data
  - Main Variables:
    - Time: trip start time, trip end time
    - Location: latitudes and longitudes of trip origins and destinations
    - People: gender, age

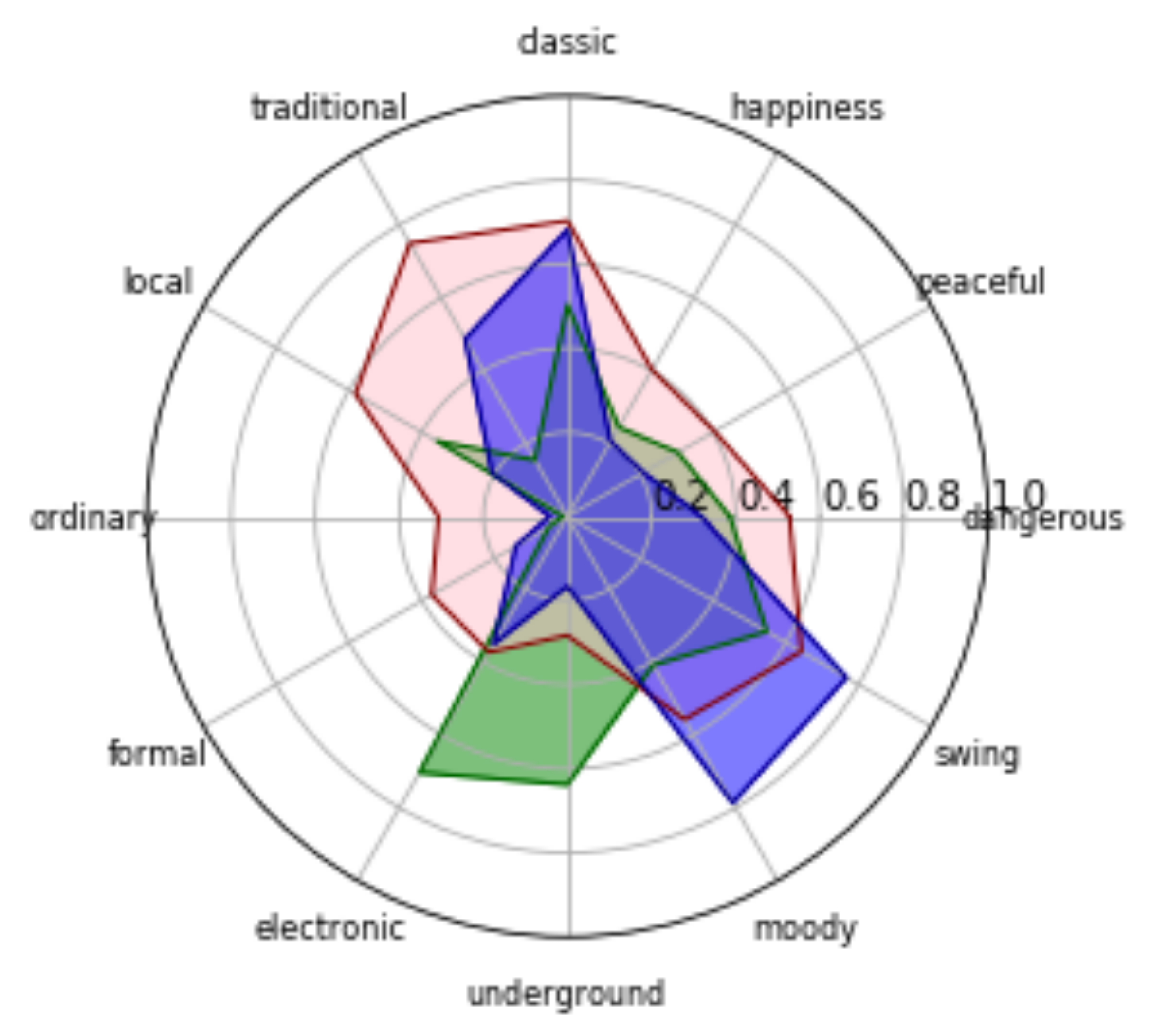
Results

- **Step 1: Styles of Some Specific Cultural Events**
- **Common Location: Millennium Park**
- Chicago House Party

  (May 27, 2017)
- Chicago Gospel Music Festival

  (June 2 & 3)
- Chicago Blues Festival

  (June 9—11)



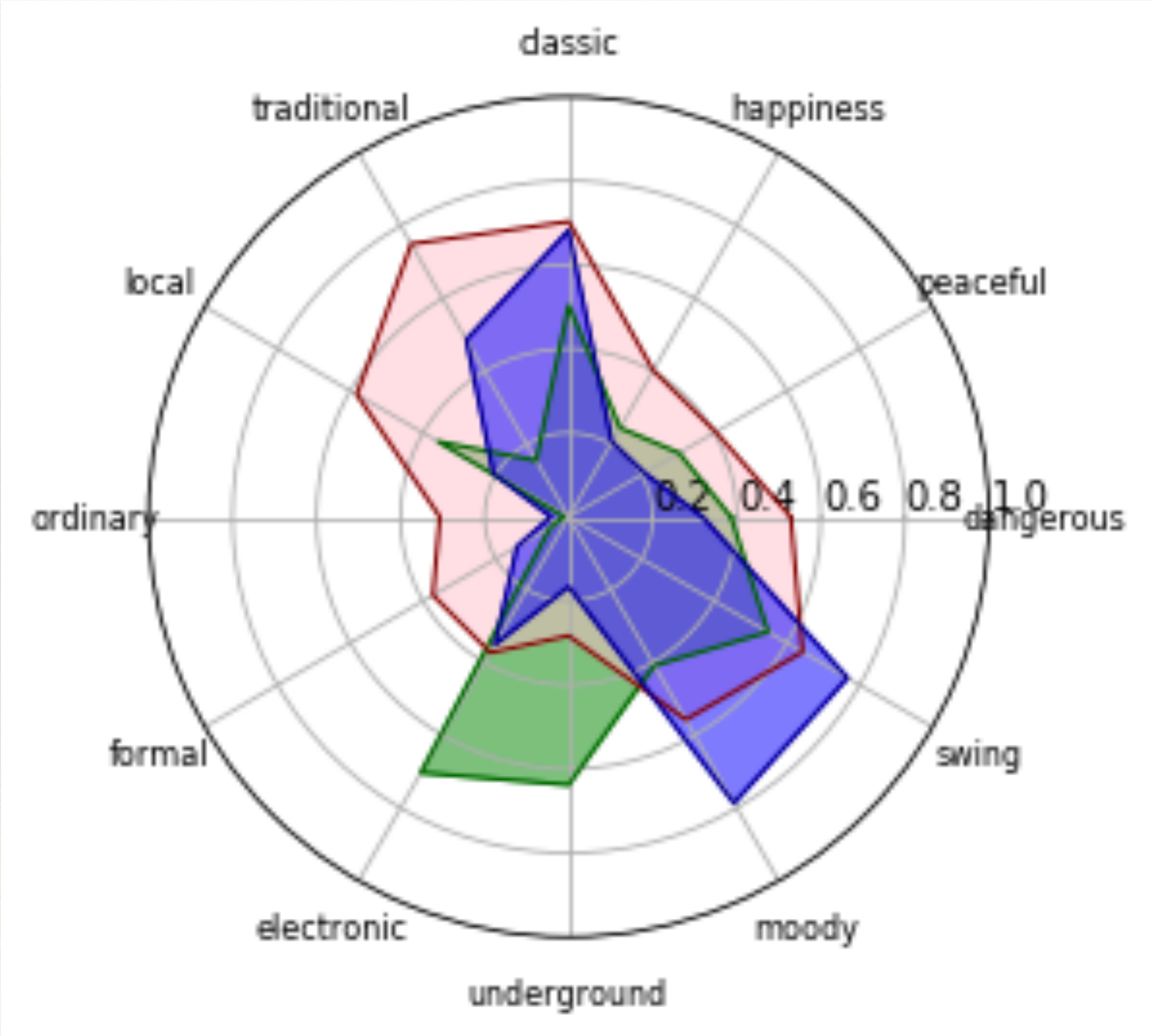
Results

- **Step 1: Styles of Some Specific Cultural Events**
- **Common Location: Millennium Park**
- Chicago House Party

  (May 27, 2017)
- Chicago Gospel Music Festival

  (June 2 & 3)
- Chicago Blues Festival

  (June 9—11)



Results

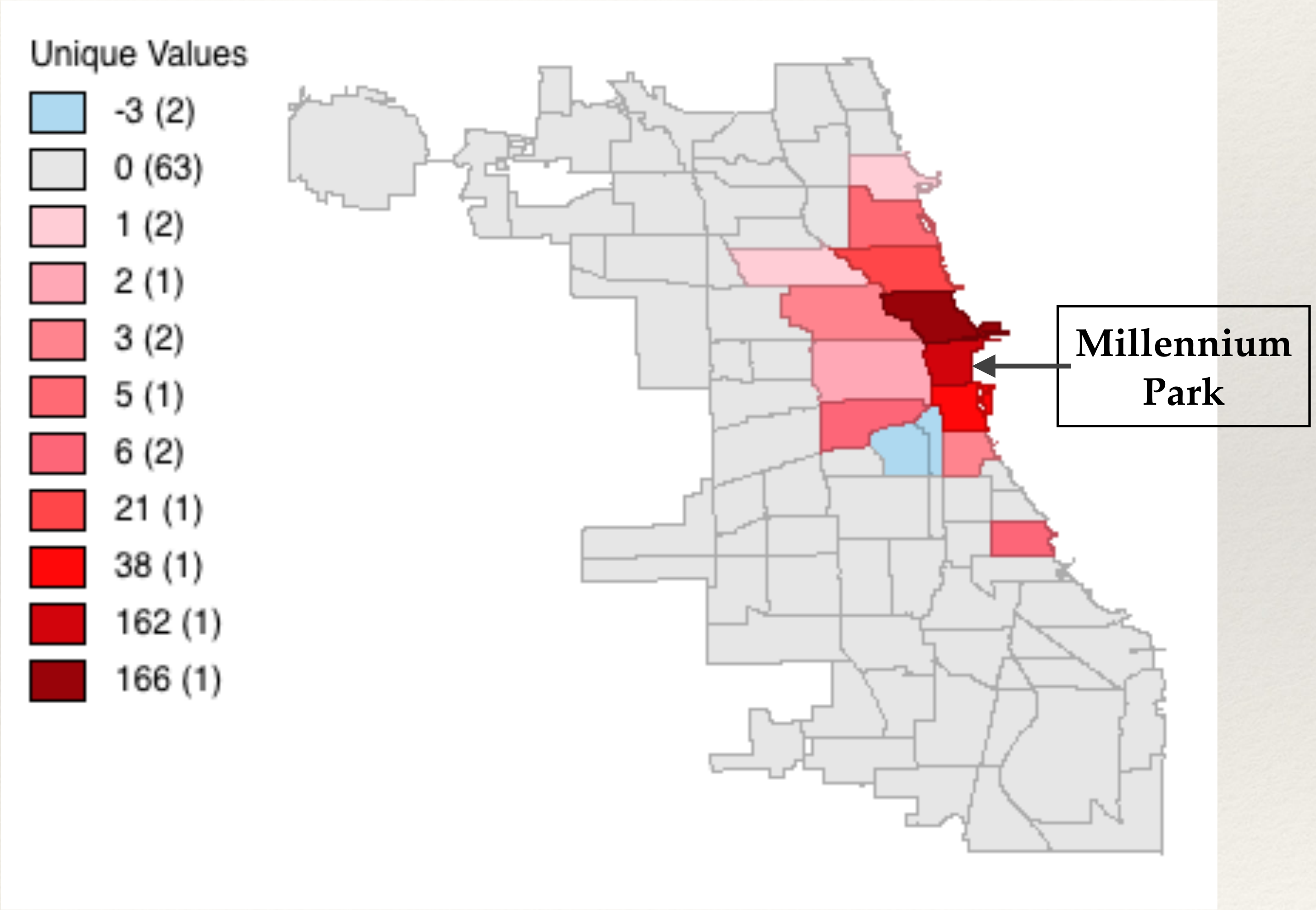
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago House Party

  (May 27)

- **Hot Zones:**
  - Near North Side (+166)
  - Loop (+162)
  - Near South Side (+38)
  - Lincoln Park (+21)

- **Cold Zones:**
  - None



Results

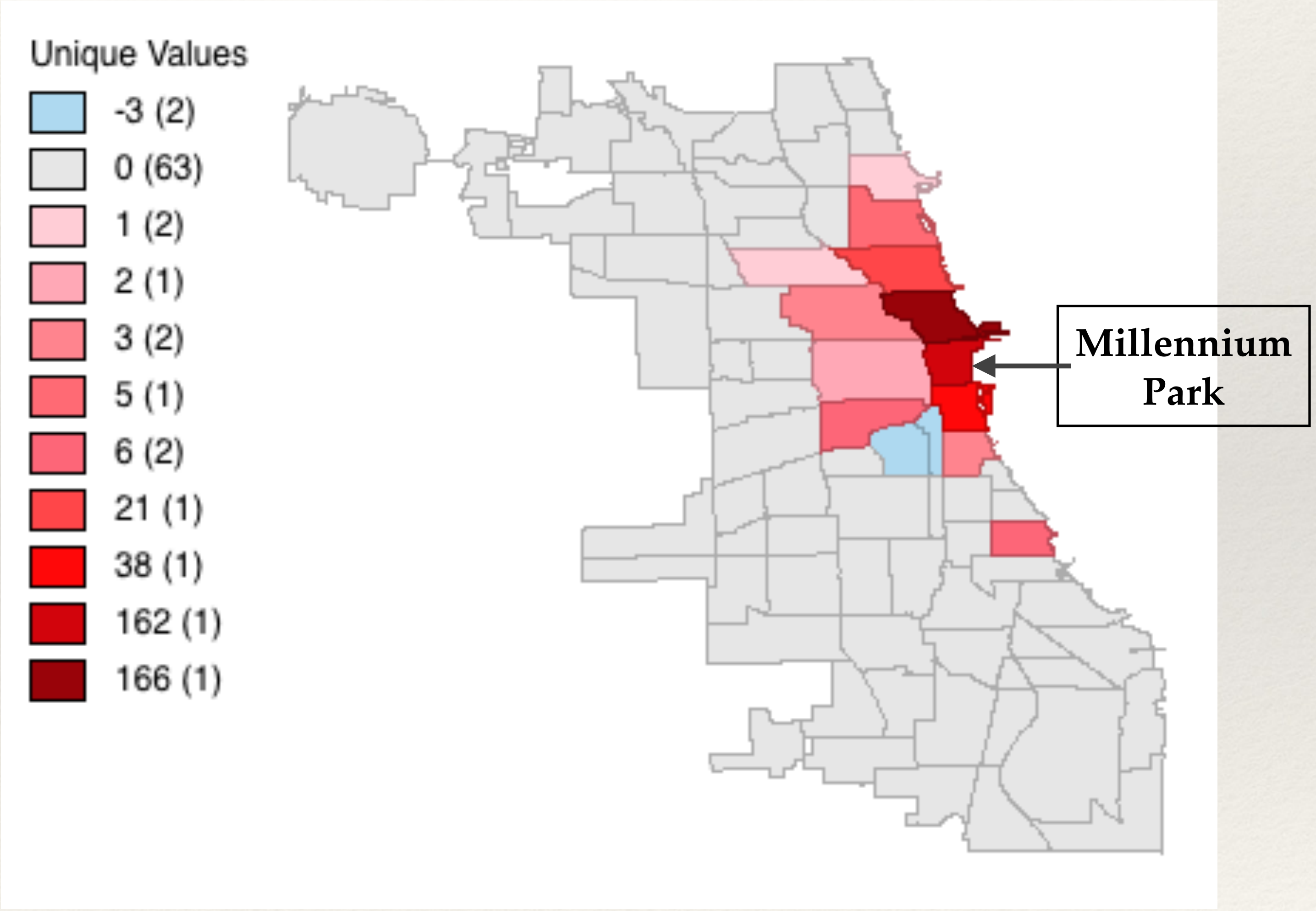
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago House Party

  (May 27)

- **Hot Zones:**
  - Near North Side (+166)
  - Loop (+162)
  - Near South Side (+38)
  - Lincoln Park (+21)

- **Cold Zones:**
  - None



Results

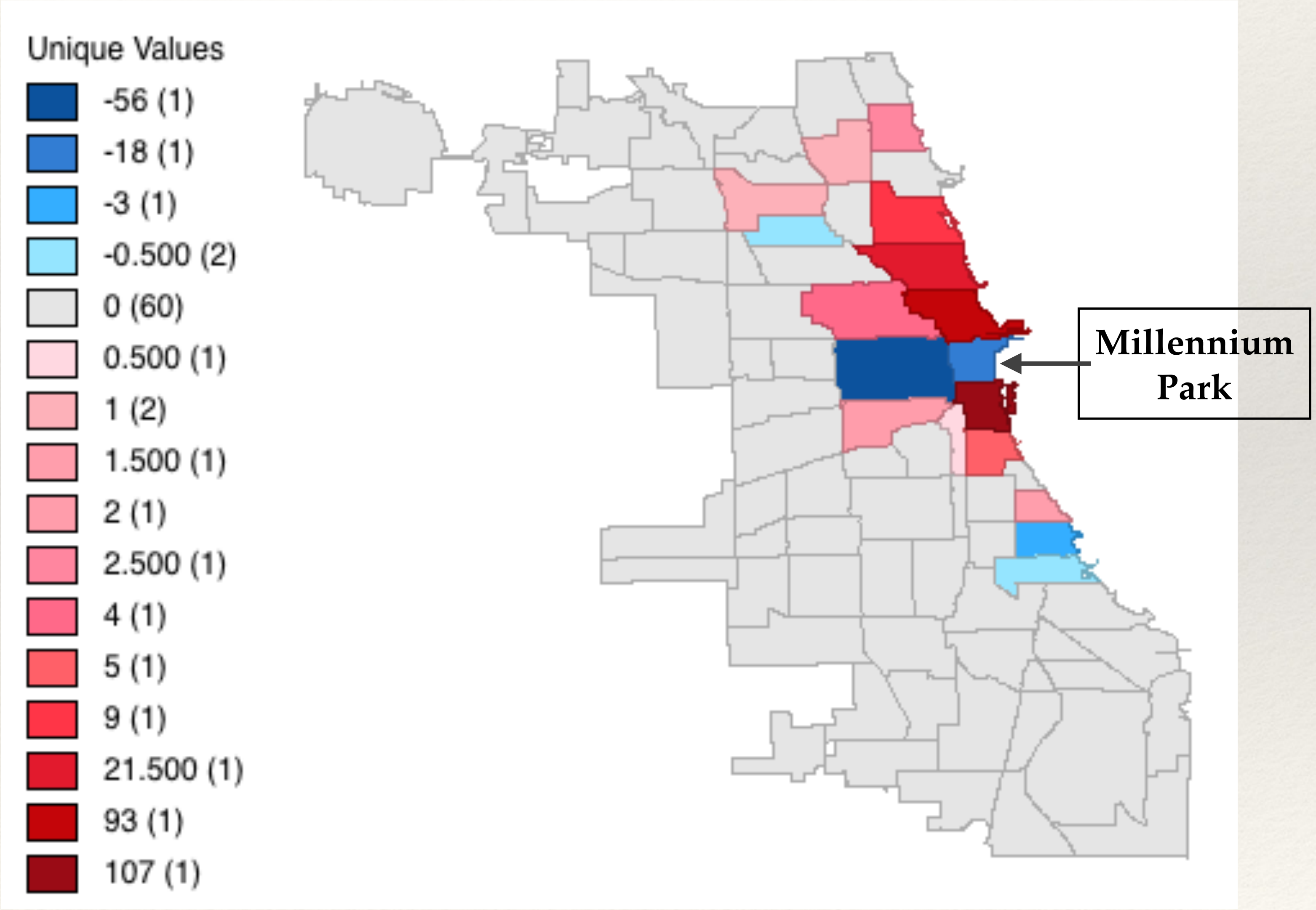
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago Gospel Music Festival

  (June 2 & 3)

- **Hot Zones:**
  - Near South Side (+107)
  - Near North Side (+93)
  - Lincoln Park (+21)

- **Cold Zones:**
  - Near West Side (-56)
  - Loop (-18)



Results

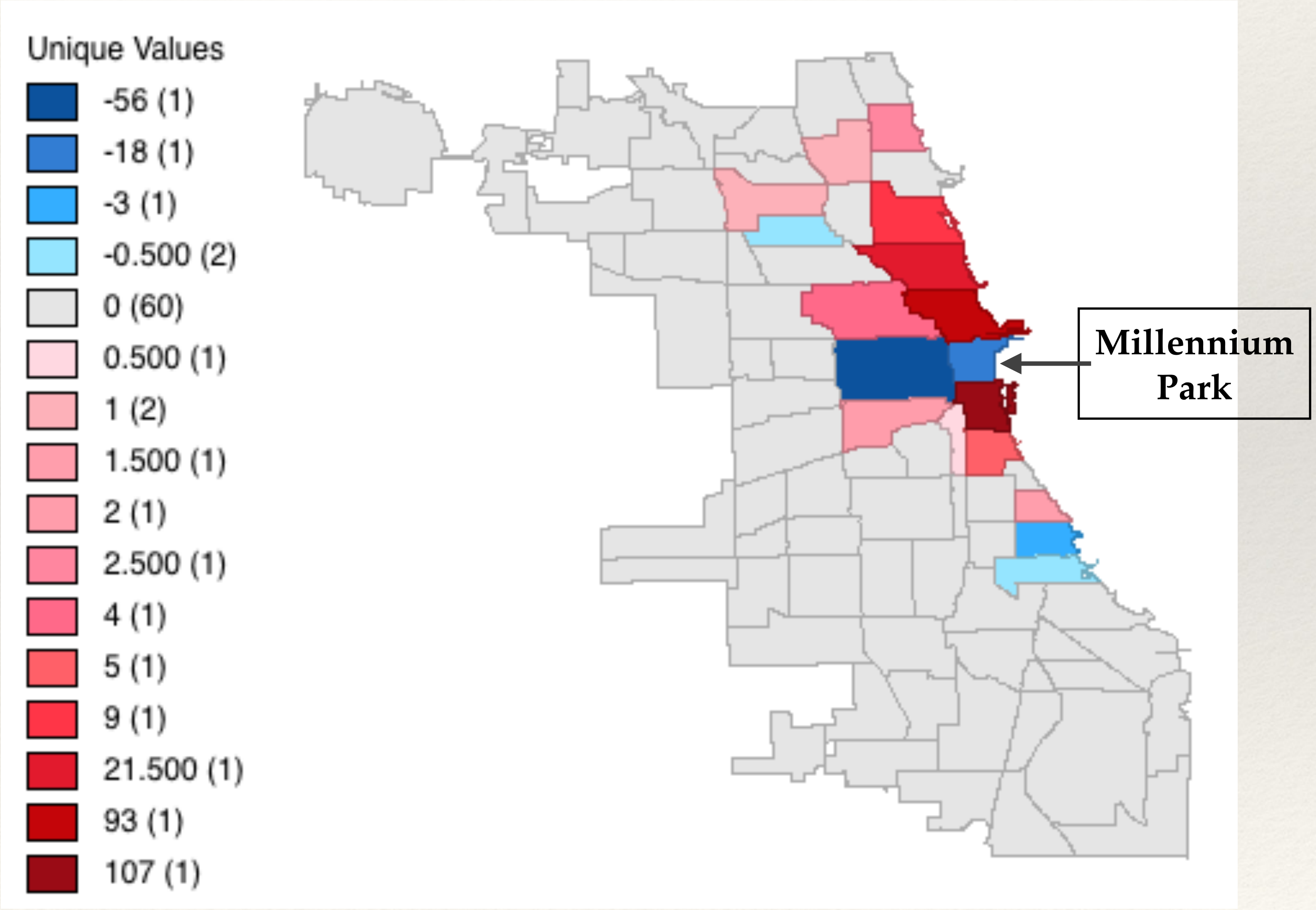
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago Gospel Music Festival

  (June 2 & 3)

- **Hot Zones:**
  - Near South Side (+107)
  - Near North Side (+93)
  - Lincoln Park (+21)

- **Cold Zones:**
  - Near West Side (-56)
  - Loop (-18)



Results

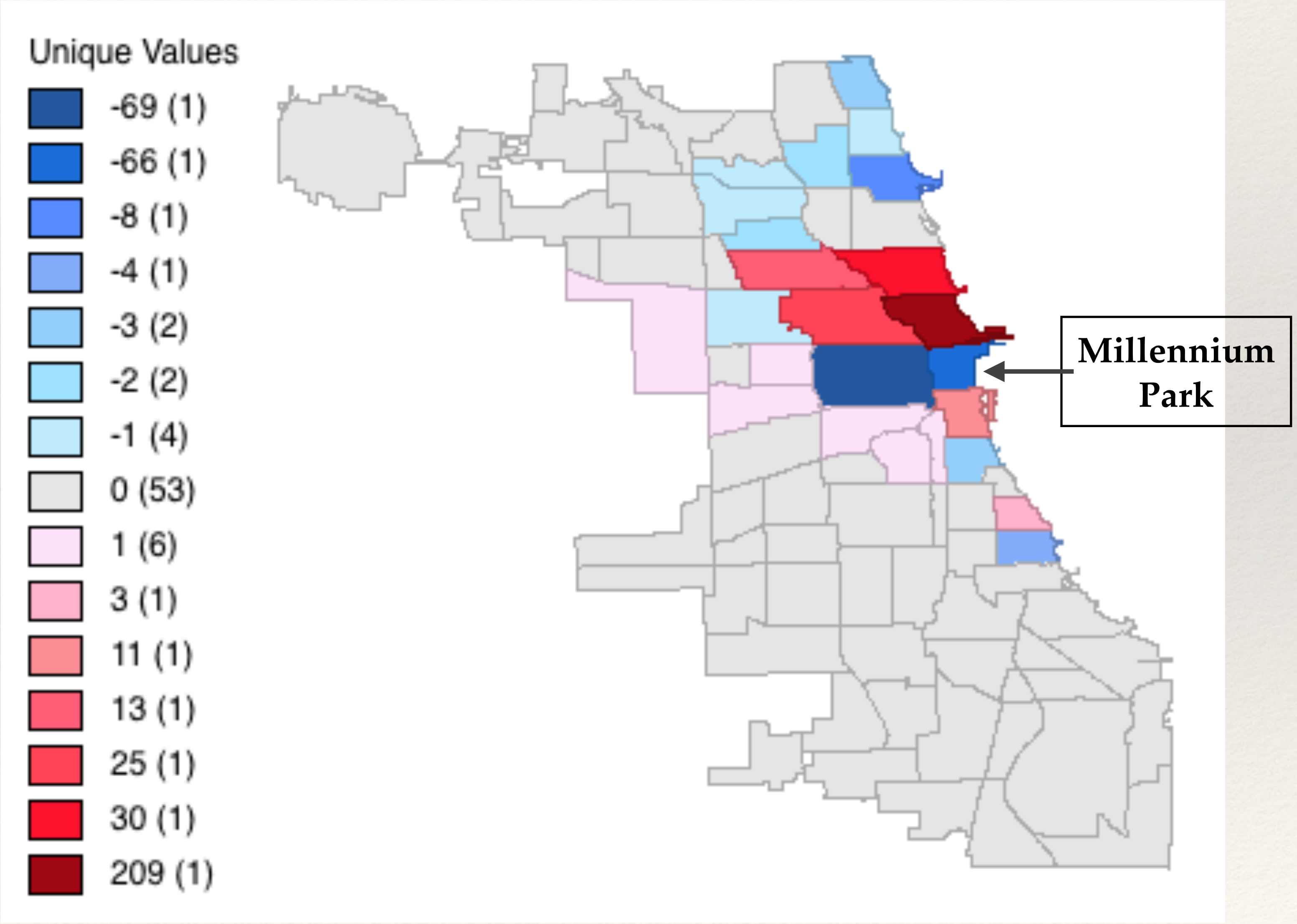
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago Blues Festival

  (June 9—11)

- **Hot Zones:**
  - Near North Side (+209)
  - Lincoln Park (+30)
  - West Town (+25)

- **Cold Zones:**
  - Near West Side (-69)
  - Loop (-66)



Results

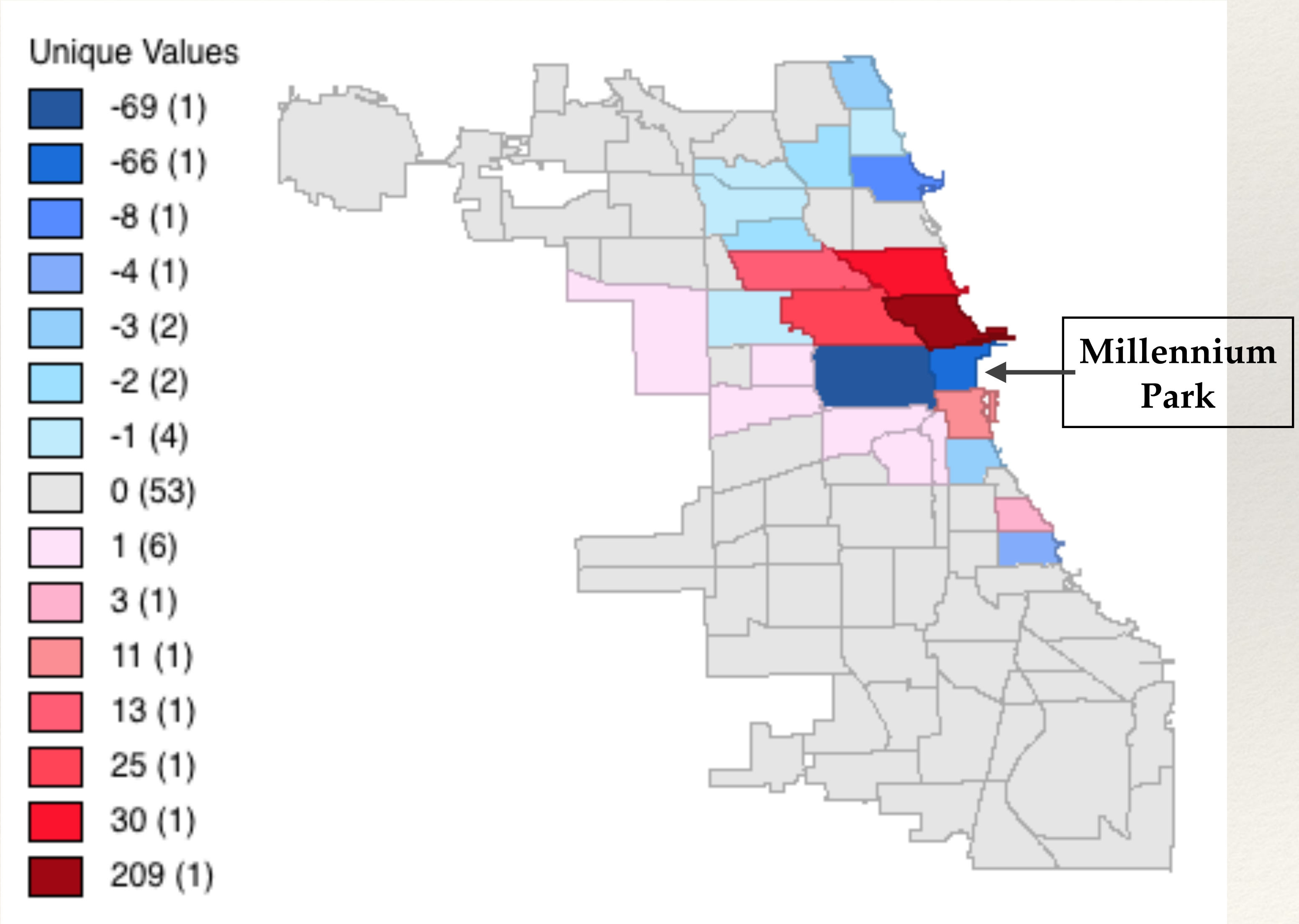
- **Step 2: Changes in transportation flows to specific cultural events**

- Chicago Blues Festival

  (June 9—11)

- **Hot Zones:**
  - Near North Side (+209)
  - Lincoln Park (+30)
  - West Town (+25)

- **Cold Zones:**
  - Near West Side (-69)
  - Loop (-66)



Problems at the Current Stage

- Weakness in proposition
- Appropriate geographic units
- Selection bias in transportation data
- Lack of demographic information of users
- Lack of socioeconomic information of users

Next Steps

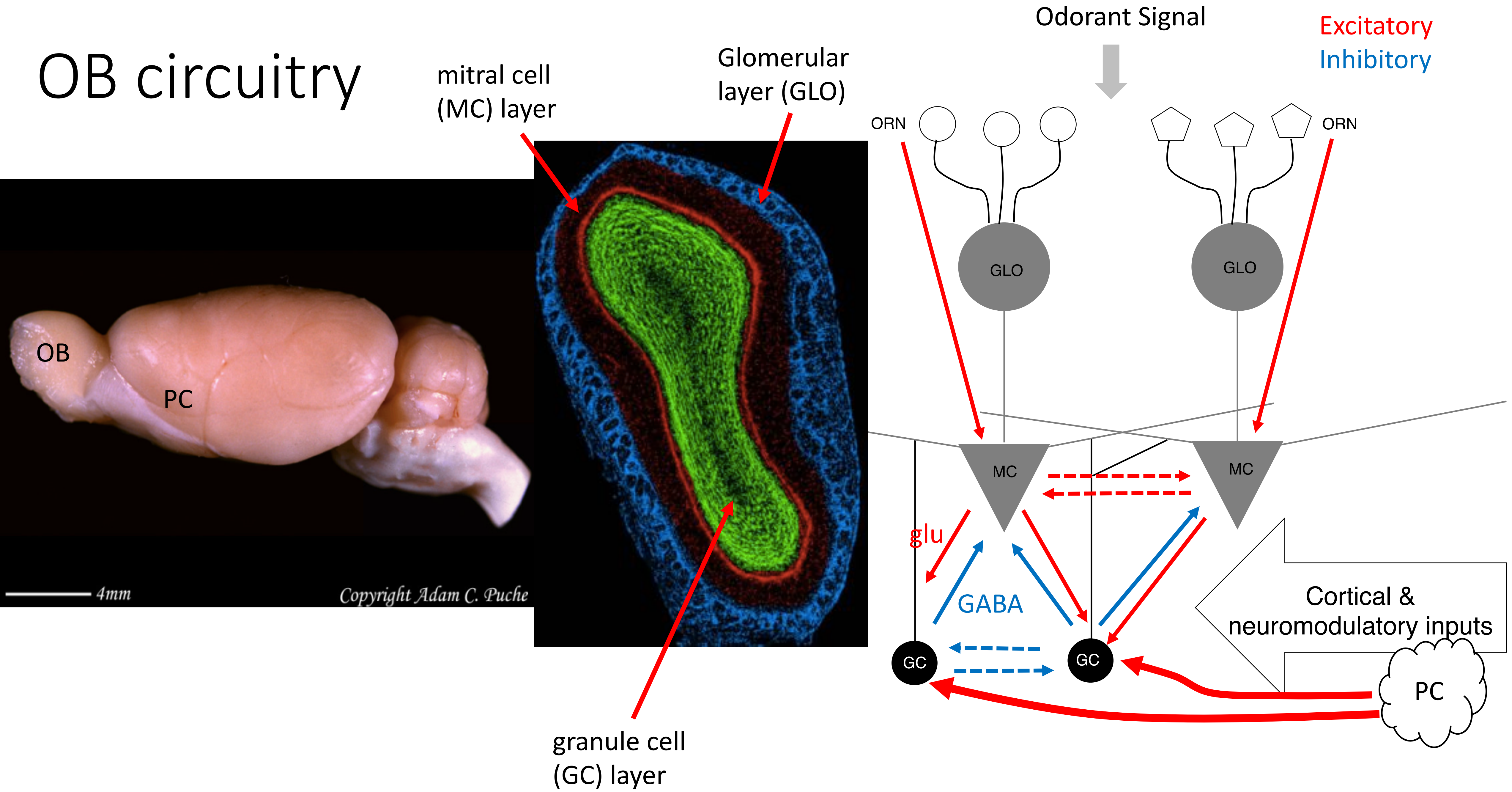
- **Data**
  - Include taxi data
  - Include more events
  - Include community data
- **Relationship between people and the cultural events**
  - Regression in counterfactual framework
- **Relationship between scenes**
  - Hierarchical linear model

Speaker Transition

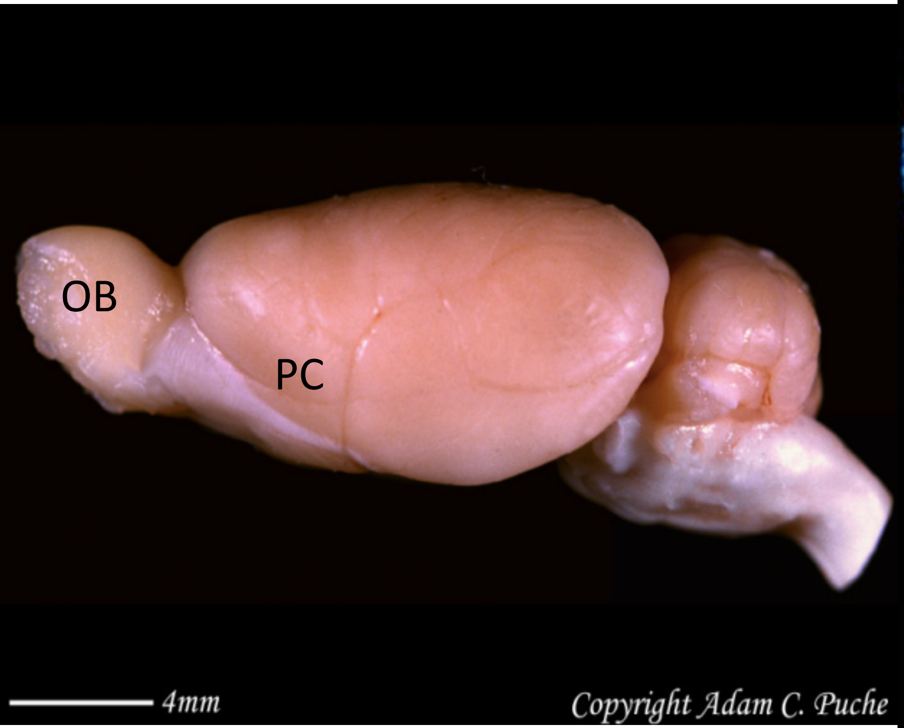
Effects of Dopamine D1 and D2 Receptors Coactivation in the Piriform Cortex (PC) on Olfactory Bulb (OB) Beta Oscillations

Wenxi Xiao
Mentor: Prof. Leslie Kay
March 29th 2018

OB circuitry



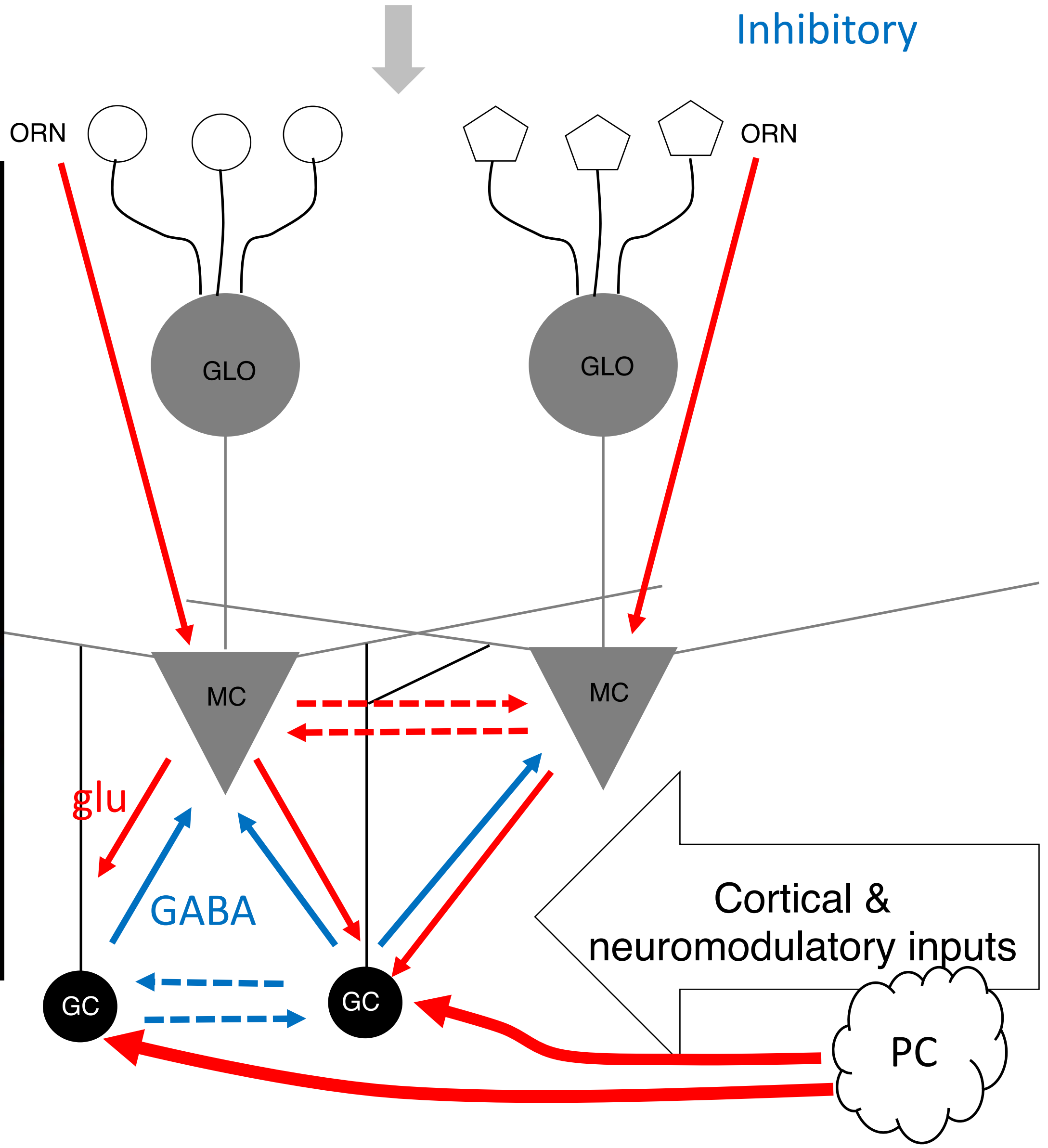
OB circuitry



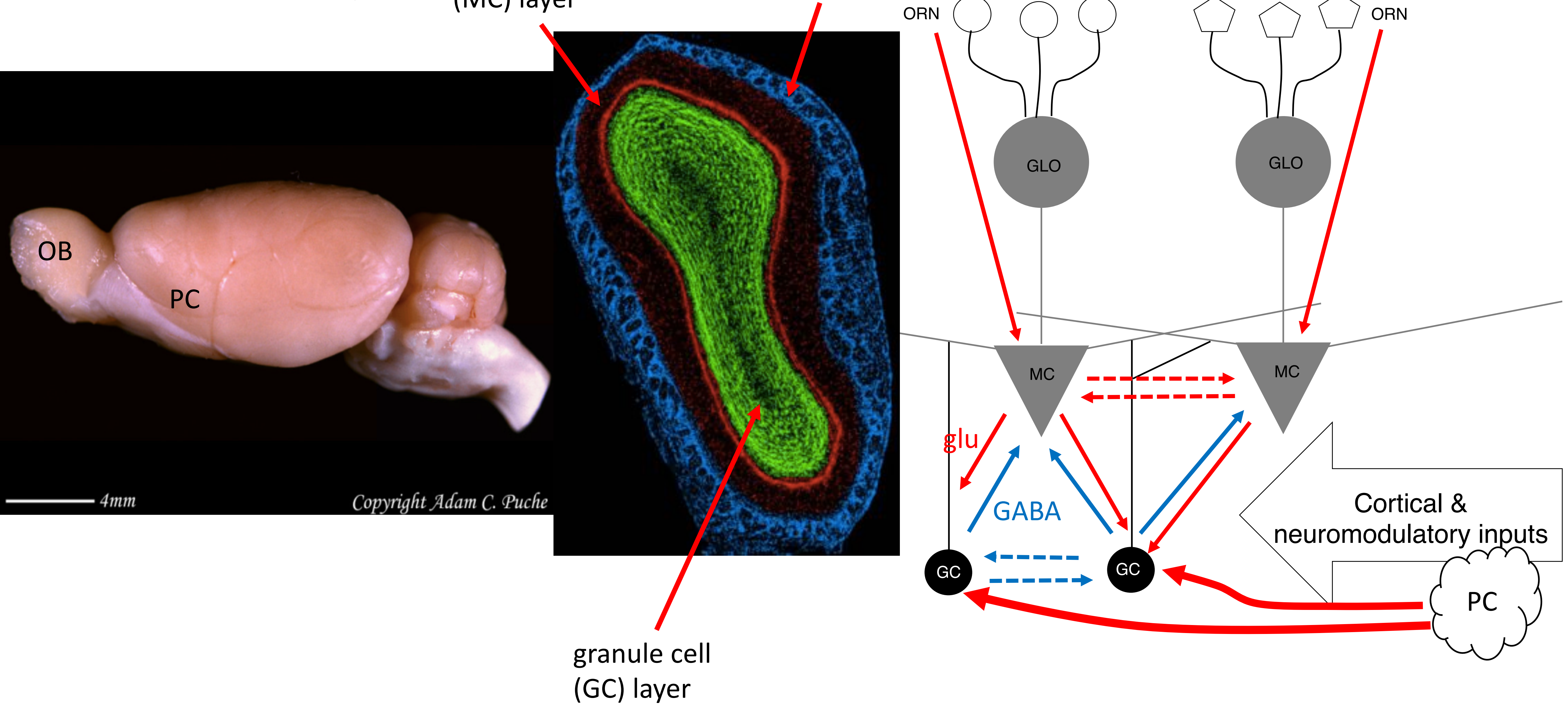
mitral cell (MC) layer

Glomerular layer (GLO)

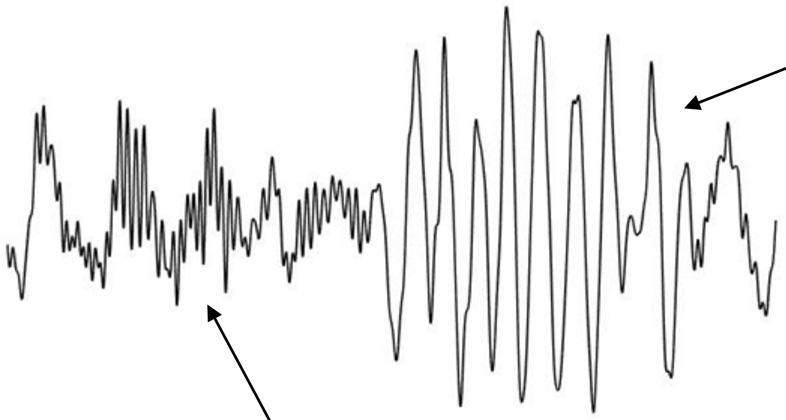
granule cell (GC) layer

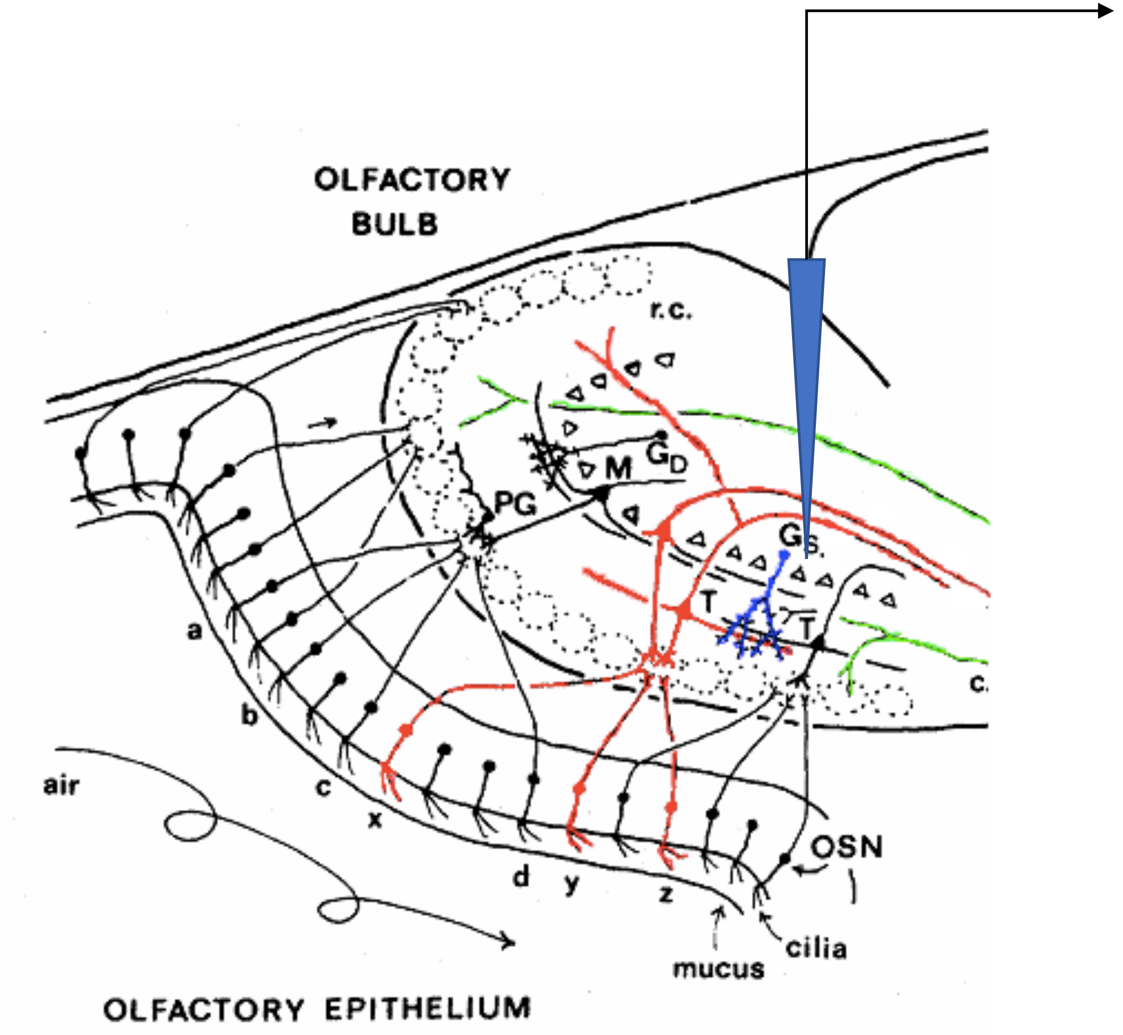


OB circuitry

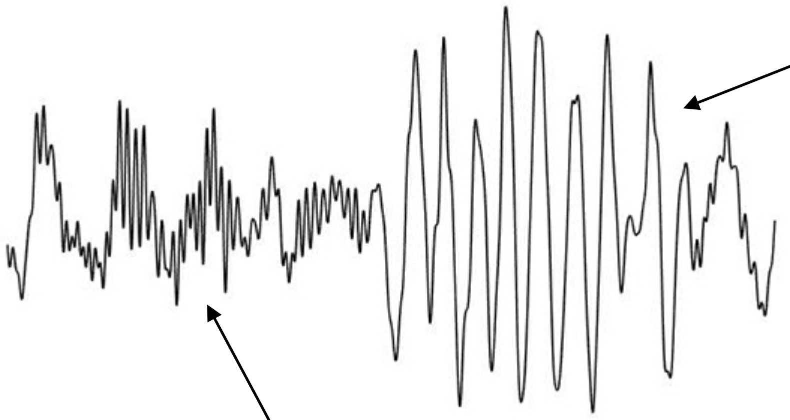


Local field potential



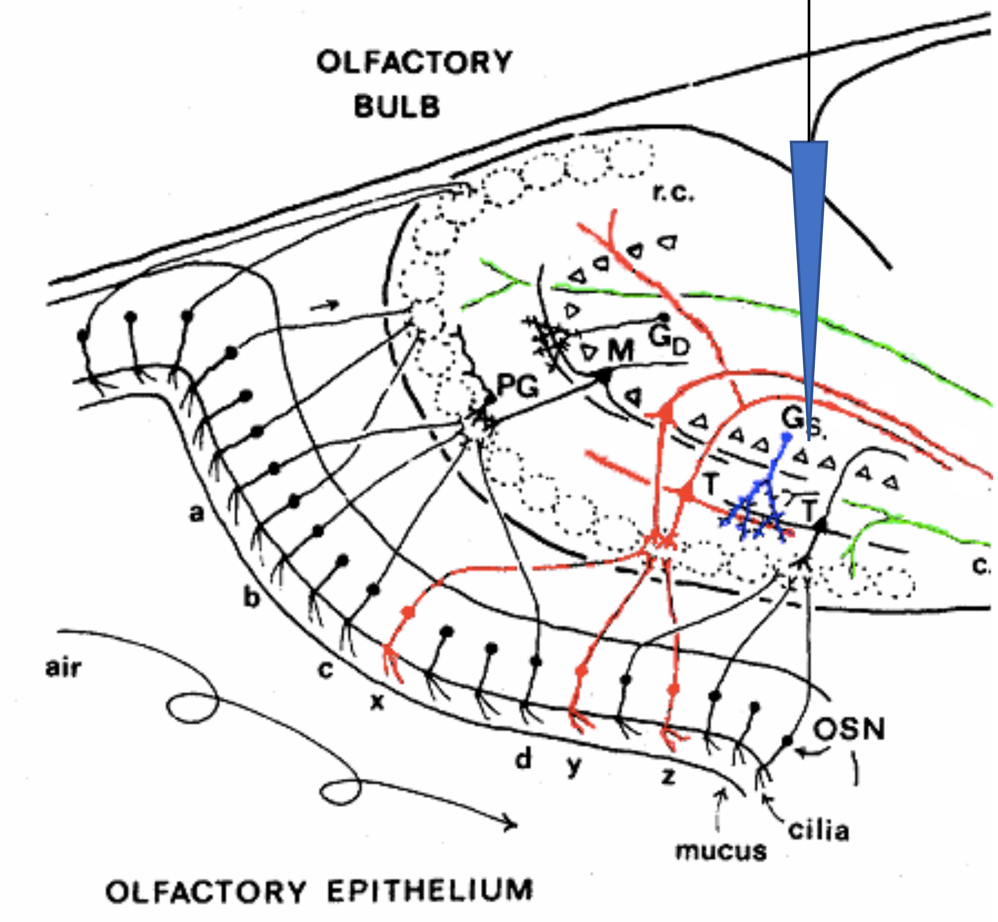


Local field potential

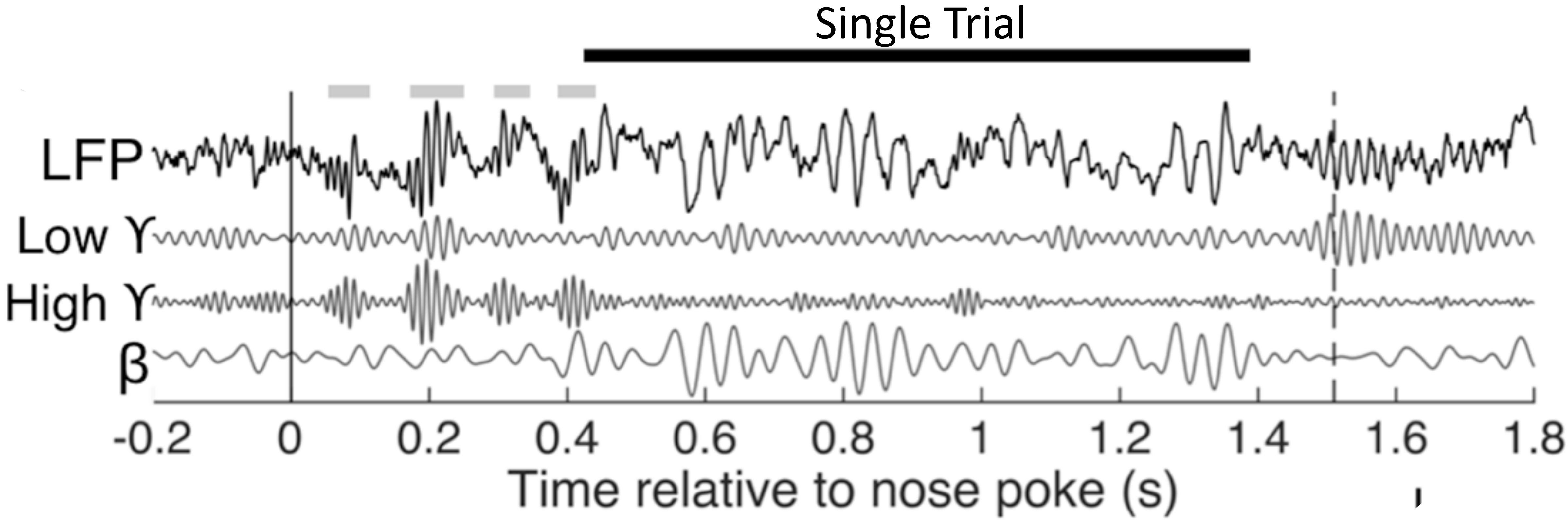


Beta (**15-30** Hz)

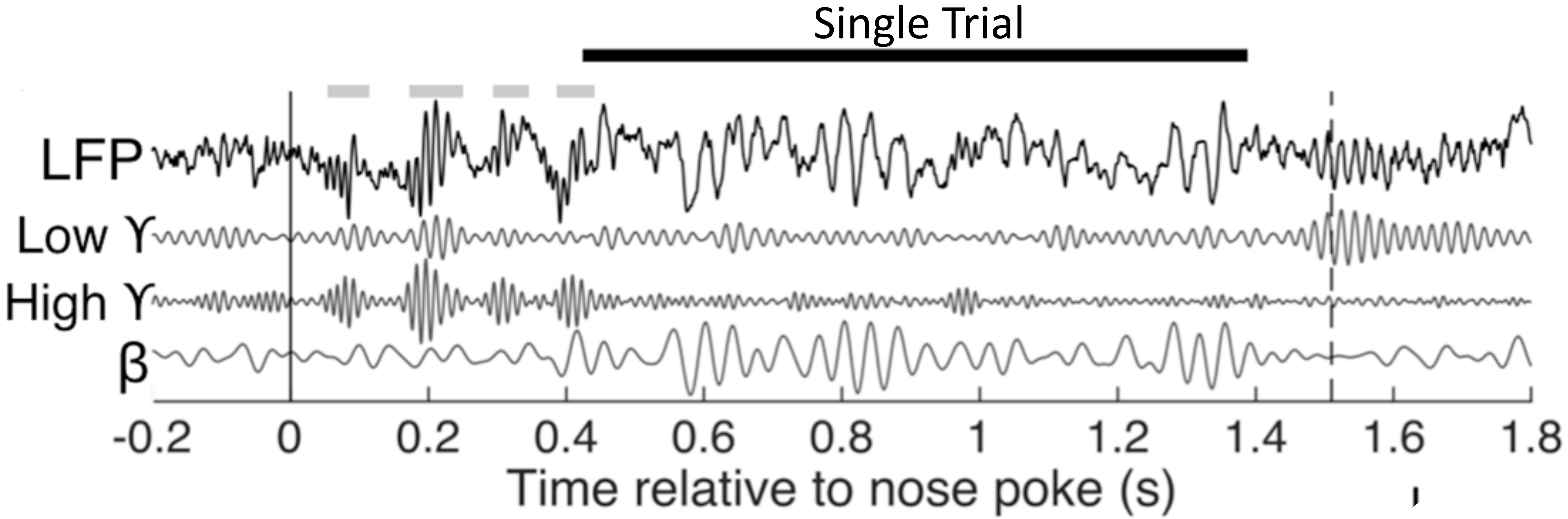
Gamma (40-100 Hz)

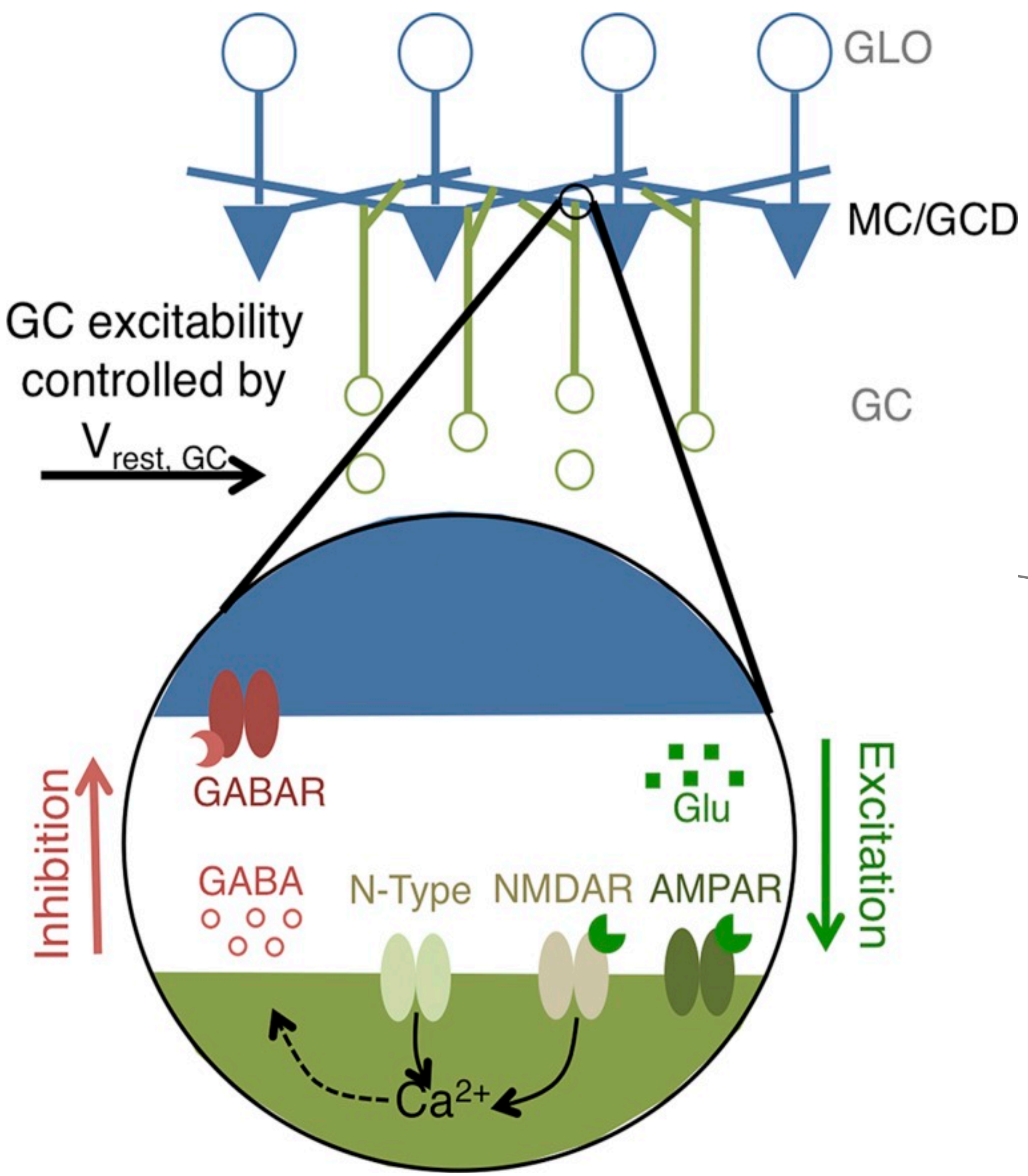


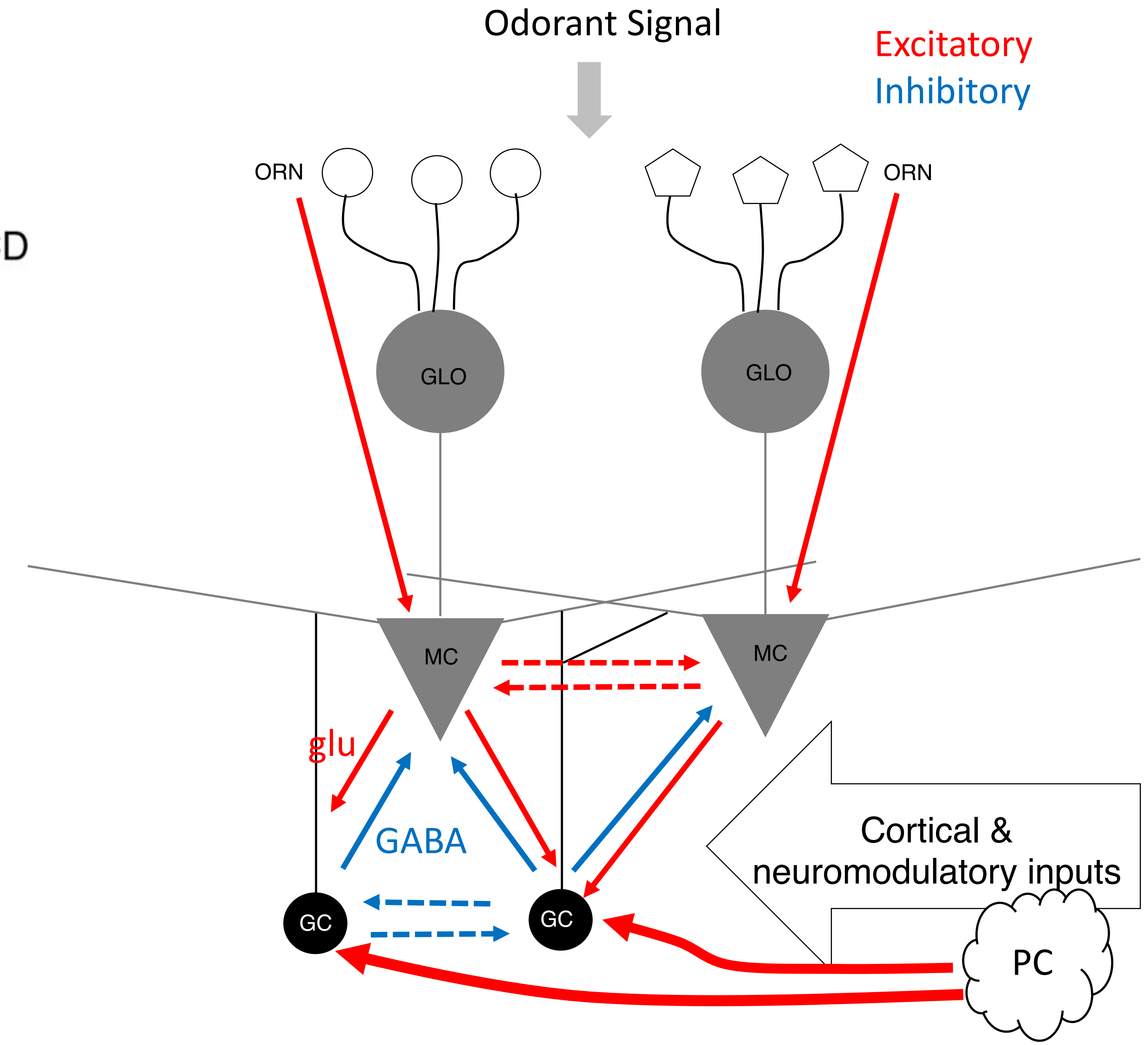
OB local field potential

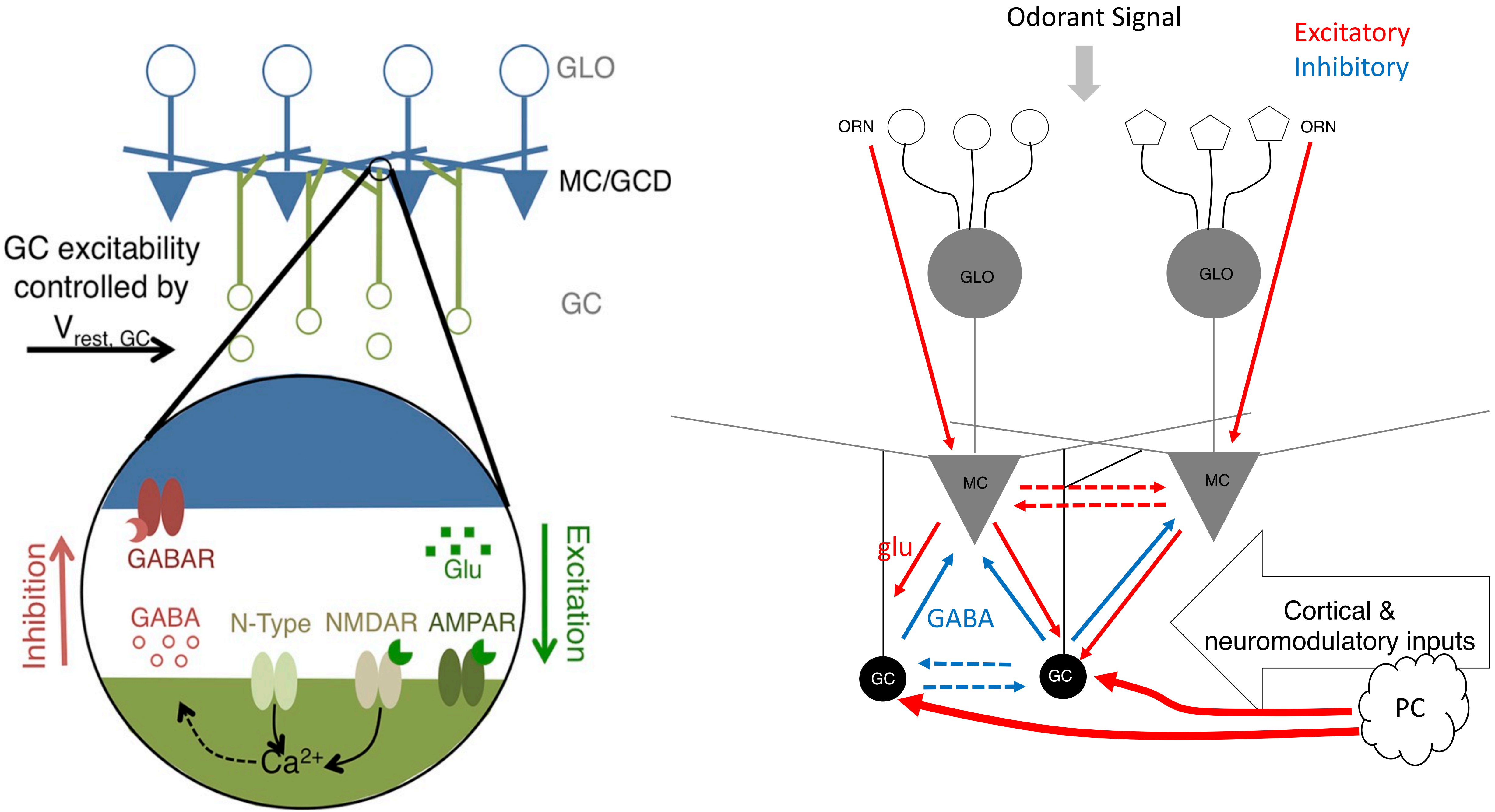


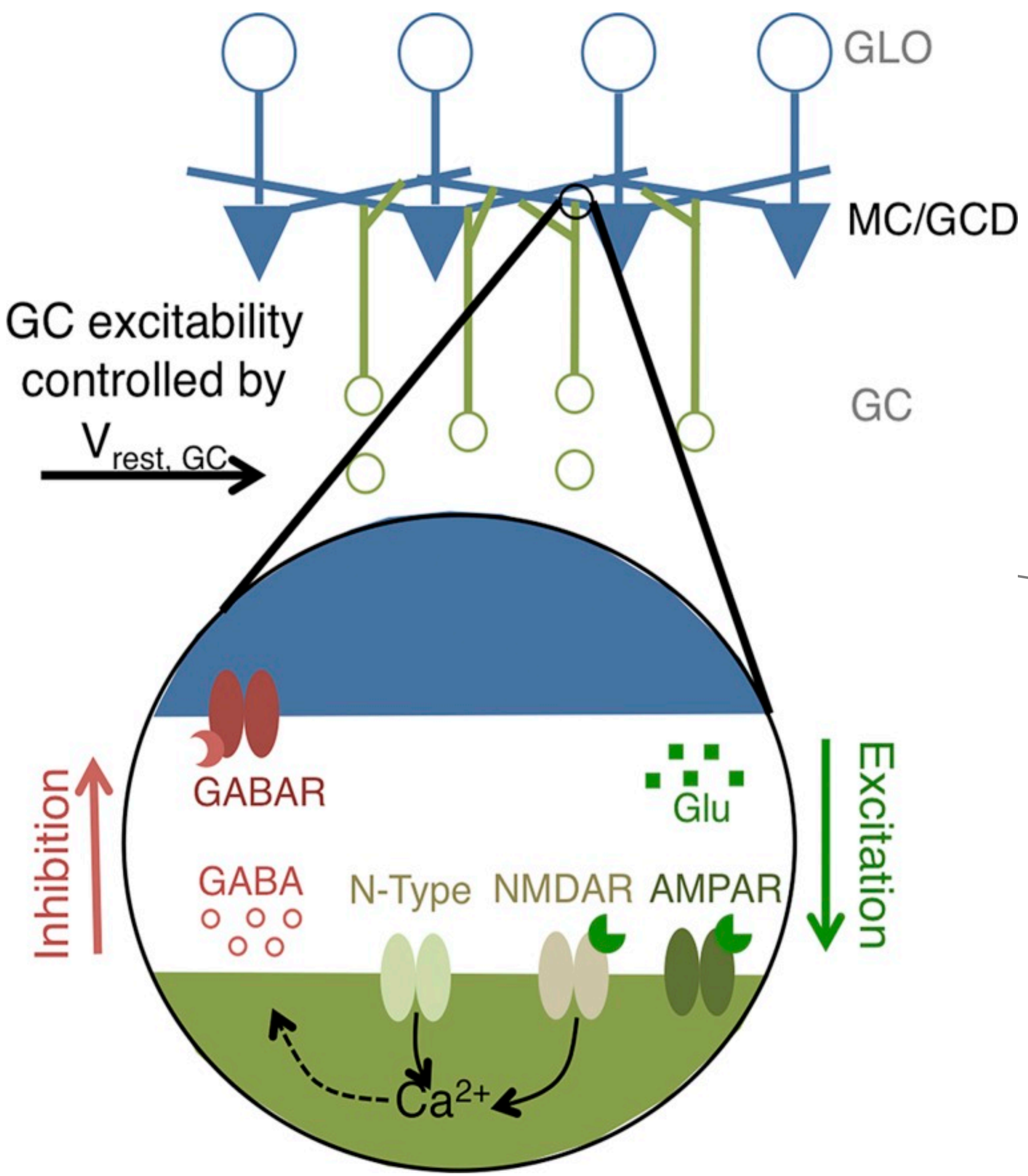
OB local field potential

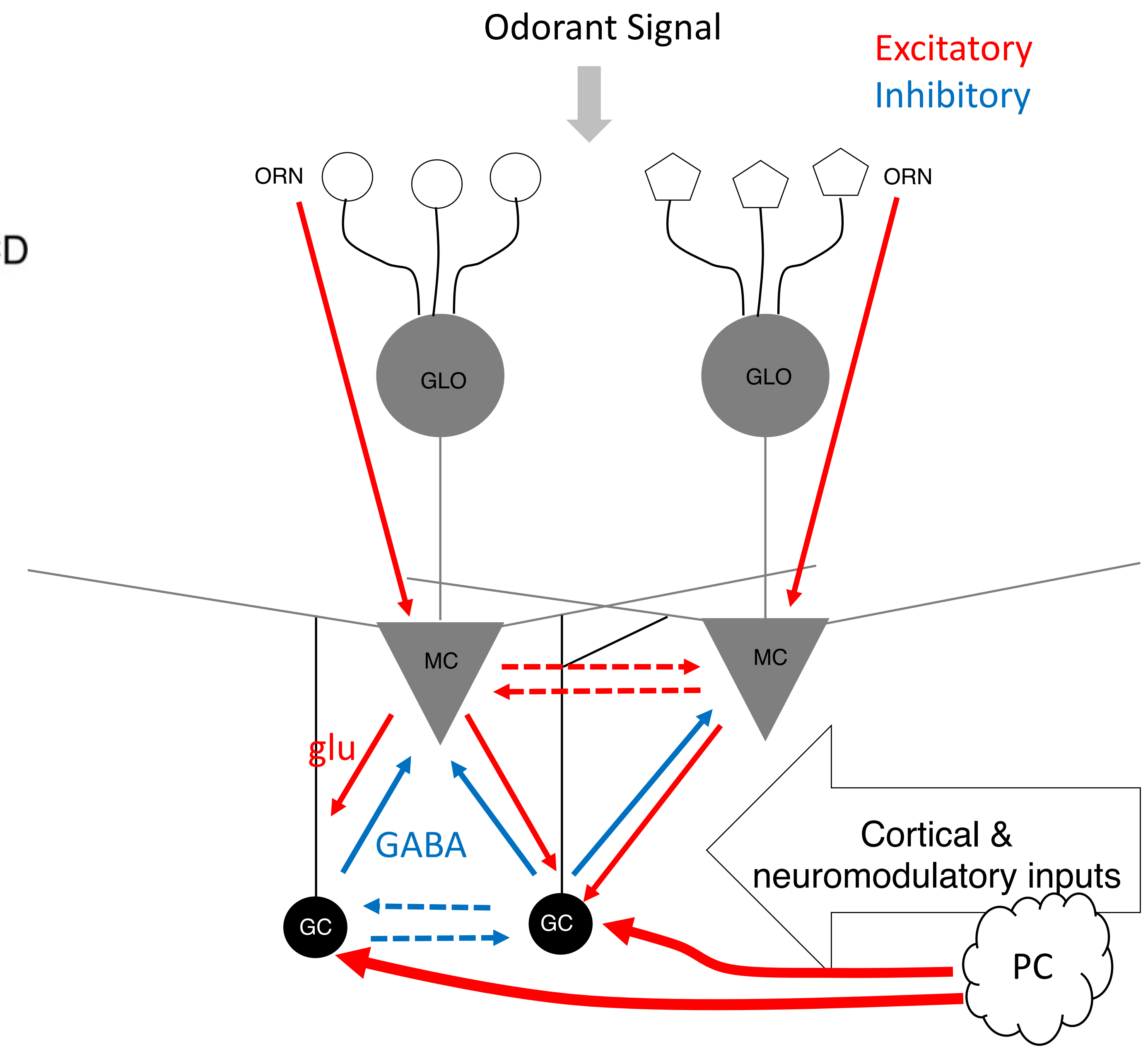



GLO
MC/GCD
GC excitability controlled by $V_{rest, GC}$
GC
GABAR
Glu
Inhibition
Excitation
GABA
N-Type
NMDAR
AMPAR
Ca^{2+}


Odorant Signal
Excitatory
Inhibitory
ORN
ORN
GLO
GLO
MC
MC
glu
GABA
GC
GC
Cortical & neuromodulatory inputs
PC

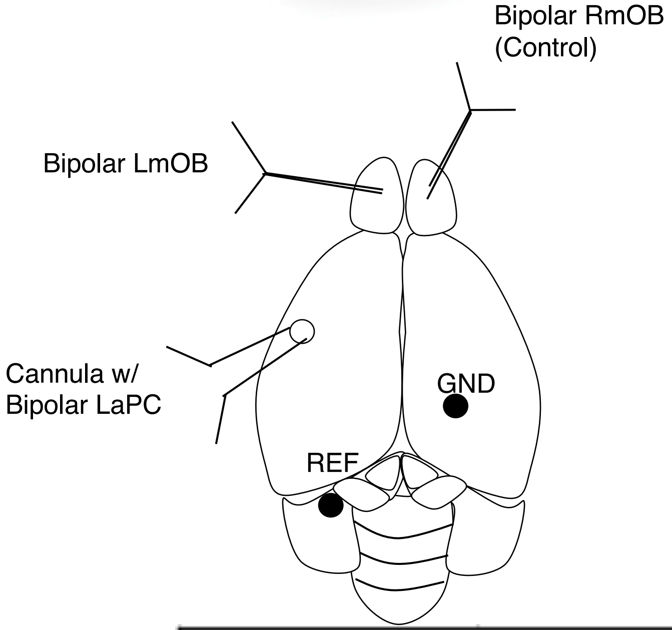

GLO
MC/GCD
GC excitability controlled by $V_{rest, GC}$
GC
GABAR
Glu
GABA
N-Type
NMDAR
AMPAR
Ca^{2+}
Inhibition
Excitation
Odorant Signal
Excitatory
Inhibitory
ORN
ORN
GLO
GLO
MC
MC
glu
GABA
GC
GC
Cortical &
neuromodulatory inputs
PC


GLO
MC/GCD
GC excitability controlled by $V_{rest,\,GC}$
GC
GABAR
Glu
Inhibition
Excitation
GABA
N-Type
NMDAR
AMPAR
Ca^{2+}

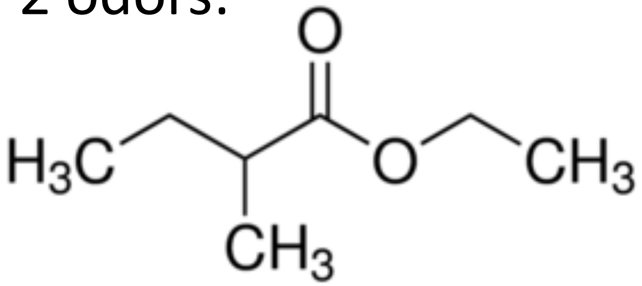

Odorant Signal
Excitatory
Inhibitory
ORN
ORN
GLO
GLO
MC
MC
glu
GABA
GC
GC
Cortical & neuromodulatory inputs
PC

How does the activity in the piriform cortex control excitability of olfactory bulb granule cells?

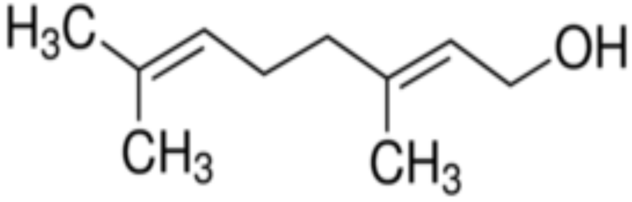
What is the involvement of the piriform cortex in gating olfactory system beta oscillations?



2 odors:



Ethyl 2-methylbutyrate (**EMB**)
high volatility: induces big beta



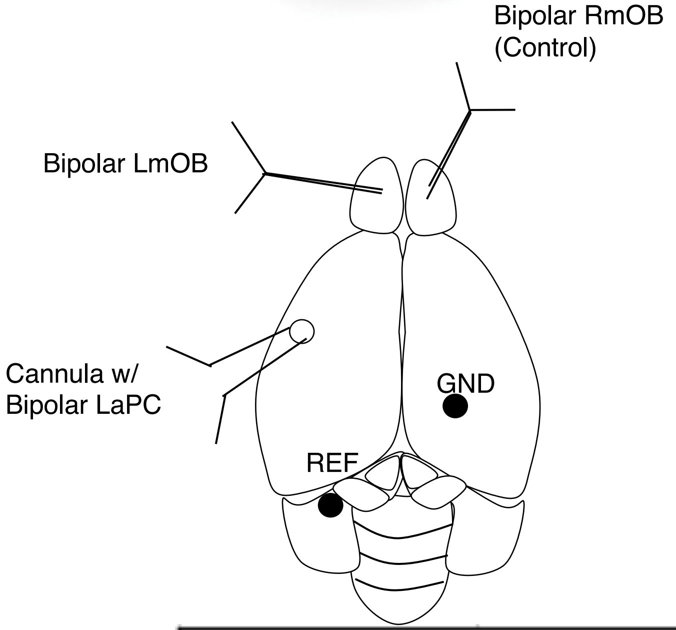
Geraniol (**Ger**)
low volatility: induces small beta

4 drugs:

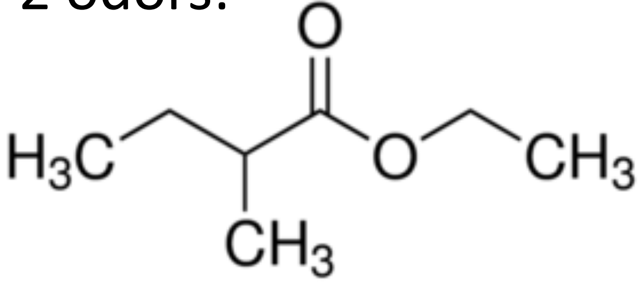
low & high dose D1/D2 agonist cocktail

low & high dose D1/D2 antagonist cocktail

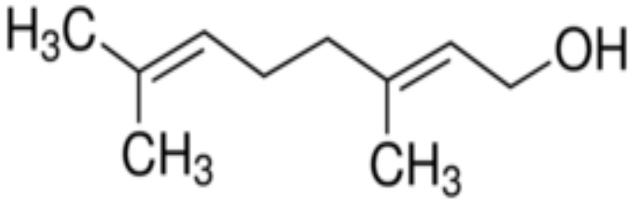
experiment timeline						
			drug/			
5 min pre behavior	12 trials odor 1 interleaved with 12 trials odor2		saline infusion	12 trials odor 1 interleaved with 12 trials odor2		10 min post



2 odors:



Ethyl 2-methylbutyrate (**EMB**)
high volatility: induces big beta



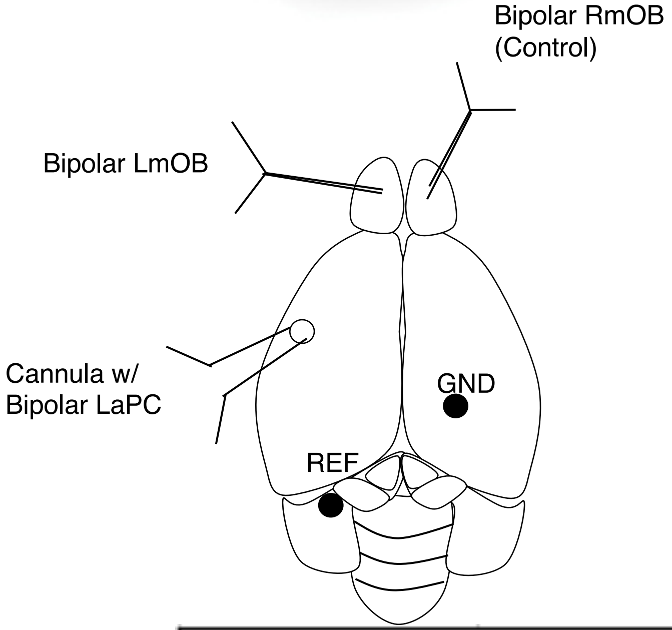
Geraniol (**Ger**)
low volatility: induces small beta

4 drugs:

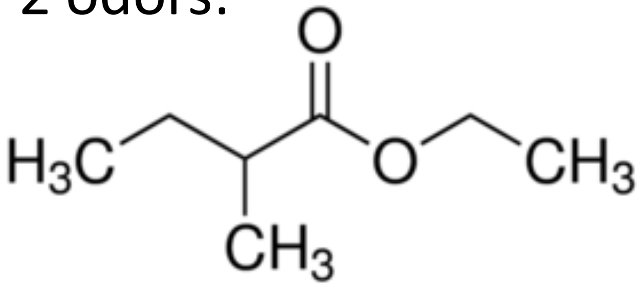
low & high dose D1/D2 agonist cocktail

low & high dose D1/D2 antagonist cocktail

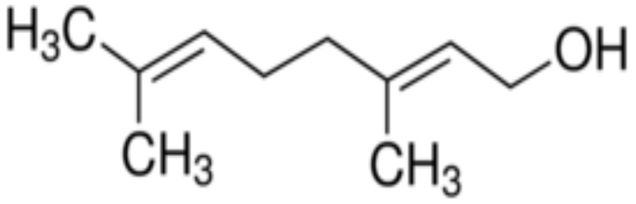
experiment timeline				
		drug/		
5 min pre behavior	12 trials odor 1 interleaved with 12 trials odor2	saline infusion	12 trials odor 1 interleaved with 12 trials odor2	10 min post



2 odors:



Ethyl 2-methylbutyrate (**EMB**)
high volatility: induces big beta



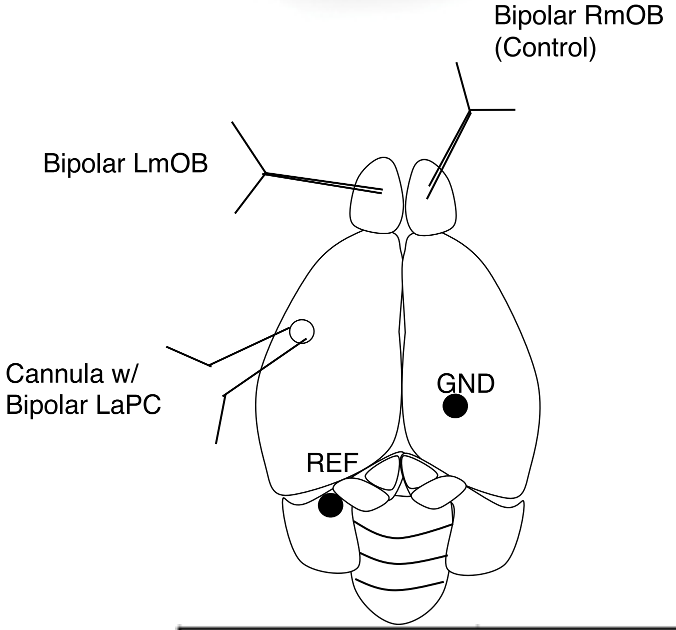
Geraniol (**Ger**)
low volatility: induces small beta

4 drugs:

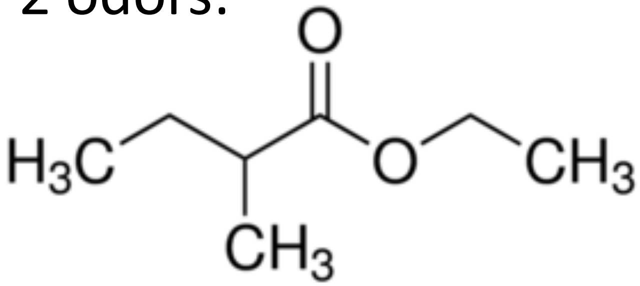
low & high dose D1/D2 agonist cocktail

low & high dose D1/D2 antagonist cocktail

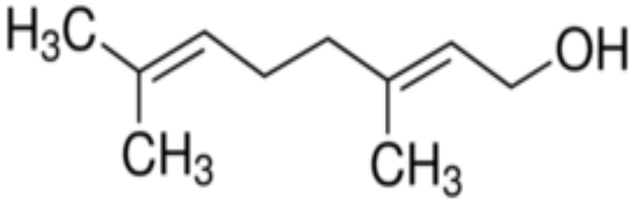
experiment timeline				
		drug/		
5 min pre behavior	12 trials odor 1 interleaved with 12 trials odor2	saline infusion	12 trials odor 1 interleaved with 12 trials odor2	10 min post



2 odors:



Ethyl 2-methylbutyrate (**EMB**)
high volatility: induces big beta



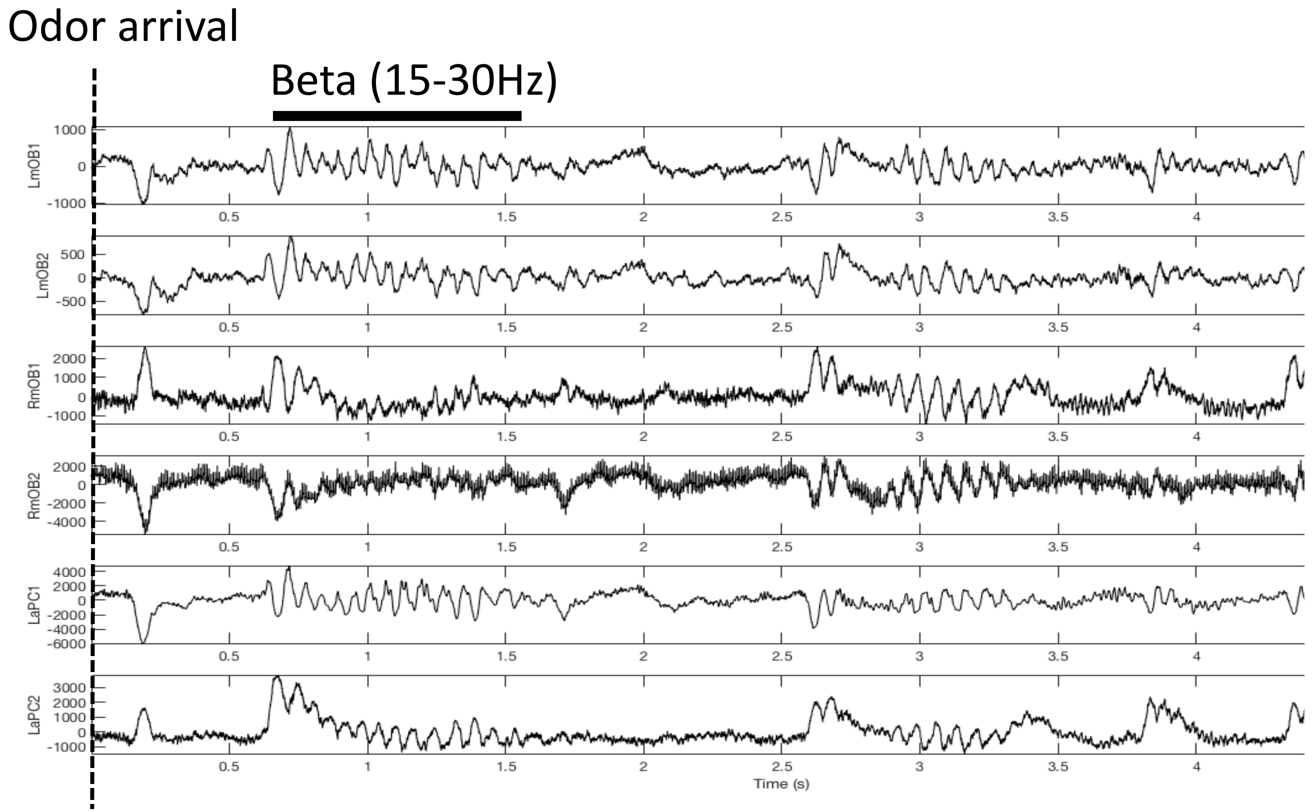
Geraniol (**Ger**)
low volatility: induces small beta

4 drugs:

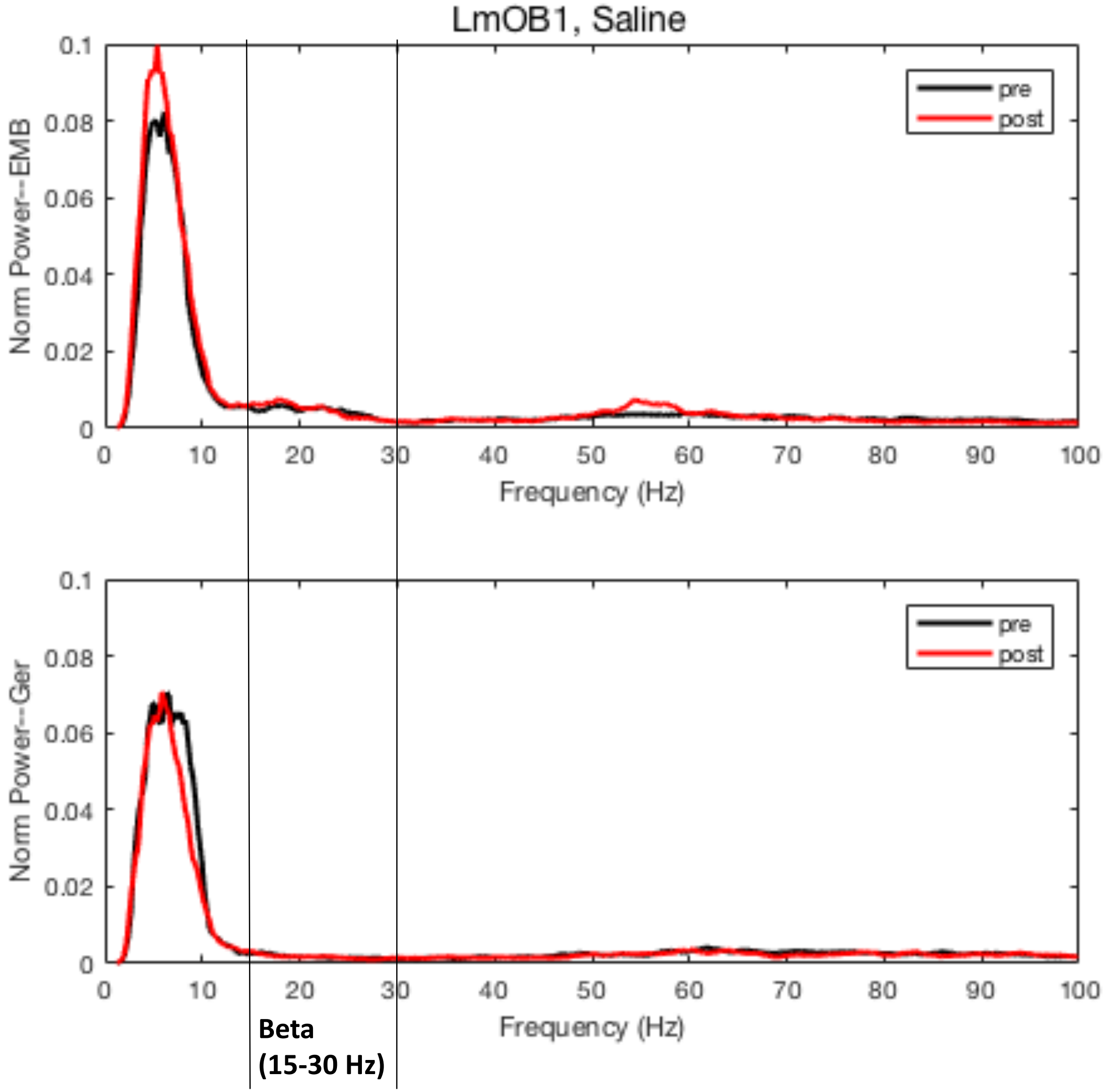
low & high dose D1/D2 agonist cocktail

low & high dose D1/D2 antagonist cocktail

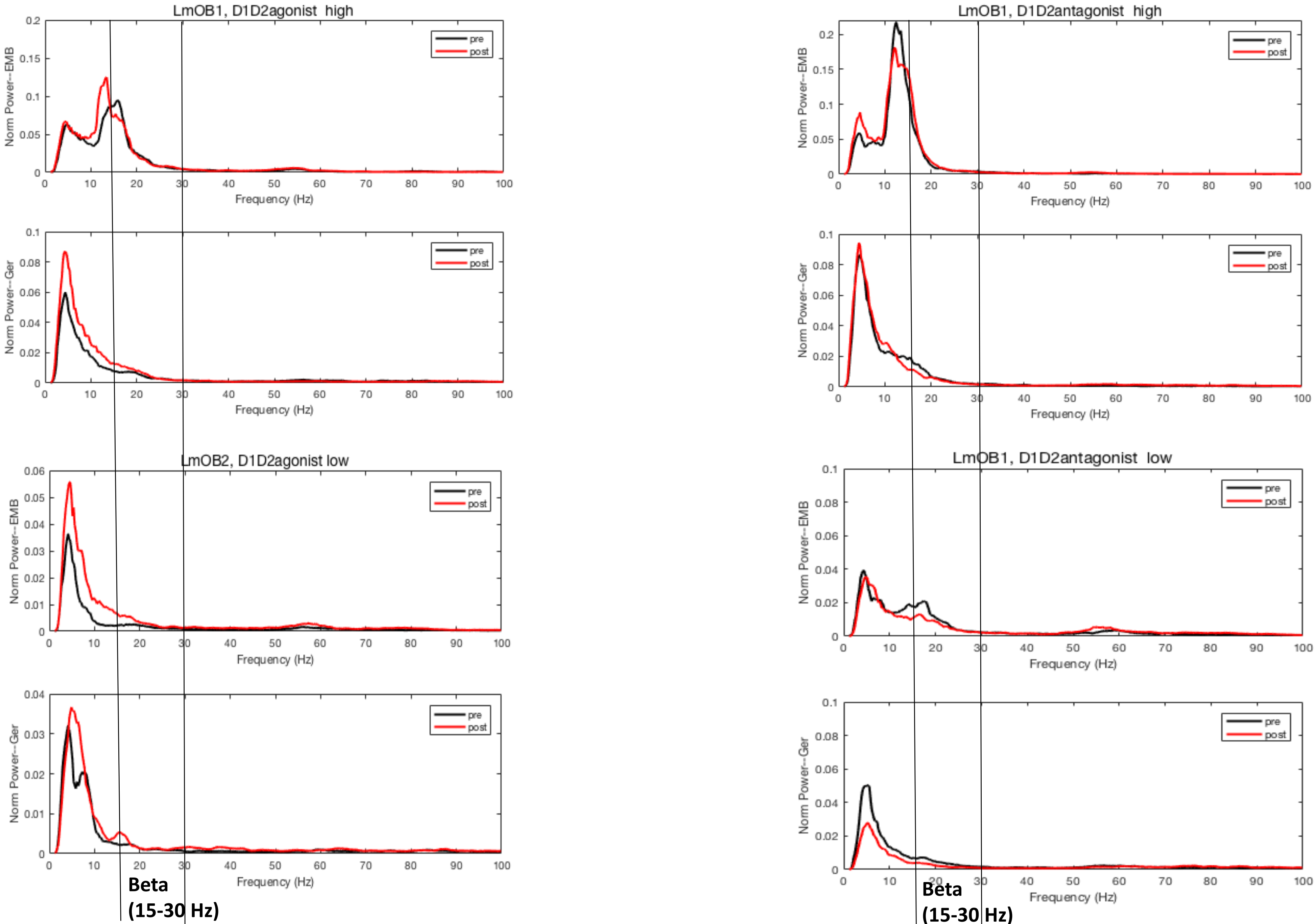
experiment timeline				
		drug/		
5 min pre behavior	12 trials odor 1 interleaved with 12 trials odor2	saline infusion	12 trials odor 1 interleaved with 12 trials odor2	10 min post



Odor arrival
Beta (15-30Hz)
LmOB1
LmOB2
RmOB1
RmOB2
LaPC1
LaPC2
Time (s)



LmOB1, Saline
Norm Power--EMB
Norm Power--Ger
pre
post
Frequency (Hz)
Beta
(15-30 Hz)


LmOB1, D1D2agonist high
LmOB1, D1D2antagonist high
LmOB2, D1D2agonist low
LmOB1, D1D2antagonist low
Norm Power--EMB
Norm Power--Ger
Frequency (Hz)
pre
post
Beta
(15-30 Hz)
Beta
(15-30 Hz)

More Analysis to be done

- Further spectral analysis of oscillations in local field potentials to examine drug duration effects (multitaper methods)
- Causal time-series analysis (i.e., Granger causality)
- 2-way ANOVA to test for effects of odors and drugs across rats on each brain region

More Analysis to be done

- Further spectral analysis of oscillations in local field potentials to examine drug duration effects (multitaper methods)
- Causal time-series analysis (i.e., Granger causality)
- 2-way ANOVA to test for effects of odors and drugs across rats on each brain region

Speaker Transition

The Impact of Chinese Parents' Absence on Their Left-Behind Children

IV vs. Machine Learning Approach

Ningyin Xu

MACSS, University of Chicago

March 29th, 2017

Outline

Background: Chinese Hukou System

Household Registration

- A "hukou" is the registered residency status of a particular individual in this system
- Each citizen was classified in an agricultural or non-agricultural hukou and further categorized by location of origin.
- Hukou status is linked to social policy.

Background: Chinese Hukou System

Household Registration

- A "hukou" is the registered residency status of a particular individual in this system
- Each citizen was classified in an agricultural or non-agricultural hukou and further categorized by location of origin.
- Hukou status is linked to social policy.

Background: Chinese Hukou System

Household Registration

- A "hukou" is the registered residency status of a particular individual in this system
- Each citizen was classified in an agricultural or non-agricultural hukou and further categorized by location of origin.
- Hukou status is linked to social policy.

Research Question

- With internal migration being more and more common, what impact do those absent parents bring to their left-behind children?
- On the one hand, higher household income results in better living/education condition.
- On the other hand, lack of parents' company might be detrimental to children.

Research Question

- With internal migration being more and more common, what impact do those absent parents bring to their left-behind children?
- On the one hand, higher household income results in better living/education condition.
- On the other hand, lack of parents' company might be detrimental to children.

Definitions/Focuses

Left-Behind Children

Left-behind children are those who age under 18 years old, and have at least one parent migrates out of the registered household area to work.

Reason for Parents' Absence

Utilizing the detailed survey our database provides, we restrict the reason for parents' absence to "migrating to work", excluding "study/military service".

Health of Left-Behind Children

Height for age, Weight for Height, and "if sick/injured in the last 4 weeks"

Education of Left-Behind Children

"whether their education levels match their ages"

Literature Review

- Rich literature on this topic comes from the perspectives of sociology and psychology.
  Liu (2007), Luo (2009), and Zhao (2013)
- Some literature focused on international migration, and proved both positive and negative impact.
  Frank and Hummer (2002), Carletto et al (2011), Antman (2012), and Schmeer (2009).
- Several literature put an emphasis on left-behind children from rural area, while neglecting those who live in urban area.
  Chen (2009), Li and Zang (2011), Brauw and Mu (2011).

Data

- China Health and Nutrition Survey (CHNS)
- 1997 - 2011, 12 provinces, 7200 households
- After cleaning, we have 5917 observations.

General Methodology

- We're focusing on the treatment effect brought by whether parents are absent. However, there might be a endogeneity problem.
- Our first hunch would be to use "Instrument Variable/Variant" and have a 2-step model.
- The question is: How to select the instrument variables?

General Methodology

- We're focusing on the treatment effect brought by whether parents are absent. However, there might be a endogeneity problem.
- Our first hunch would be to use "Instrument Variable/Variant" and have a 2-step model.
- The question is: How to select the instrument variables?

Classic Approach

- From previous literature, one classic instrument variable for "internal migration" is "the ratio of internal migration in the neighborhood", since it would influence parents' decision but not the condition of their left-behind children.
- Other literature have also used the education level of parents.

Machine Learning Approach

- We can also consider this as a "model selection" problem in machine learning, thus we plan to use machine learning approach to find a set of variables that could be used as our "instrument variant".
- Still in progress. (Feel free to give me some advice about this part!)

Descriptive Statistics (for dependent variables)

Table: Parents don't migrate

	Rural Area		Urban Area	
	Mean	Var.	Mean	Var.
get sick/injured in the past 4 weeks	6.92%	0.06	8.51%	0.08
female	46.19%	0.25	46.11%	0.25
age	10.00	21.39	10.26	22.62
weight	31.62	222.65	35.88	282.21
height	131.29	726.01	136.34	796.56
education (for 6-18 years old)	37.63%	0.23	52.33%	0.25

Descriptive Statistics (for dependent variables)

Table: At least one parent migrates

	Rural Area		Urban Area	
	Mean	Var.	Mean	Var.
get sick/injured	8.93%(+)	0.08	6.35%(-)	0.06
female	42.96%(-)	0.25	47.65%(+)	0.25
age	8.96(-)	21.84	9.25(-)	19.08
weight	28.49(-)	207.47	31.74(-)	294.75
height	125.63(-)	760.67	130.12(-)	694.17
education	40.00%(+)	0.24	50.00%(-)	0.26

Model

a 2SLS model:

$$y_i = \begin{cases} 1 & y_i^* < \mu_1 \\ 2 & \mu_1 < y_i^* < \mu_2 \\ \cdots N & y_i^* > \mu_{J-1} \end{cases}$$

where:

$$y_i^* = \alpha_0 + X_i\alpha_1 + T_i\alpha_2 + \epsilon_i$$

where T_i is the treatment variable.

$$T_i = \begin{cases} 1 & X_i\beta_1 + Z_i\beta_2 + u_i > 0 \\ 2 & otherwise \end{cases}$$

Regression results for treatment model

Table: Results of Endogenous Treatment Variable Model (y: whether parents migrate)

	All obs.	Age 0-6	Age 6-12	Age 12-18	Rural	Urban
migration rate	7.390***	6.595***	8.103***	7.503***	7.384***	7.319***
gender	-0.019	0.032	-0.022	-0.042	-0.033	0.062
age	-0.022***	0.002	-0.050**	-0.040**	-0.023***	-0.016
hhincome	-0.000***	-0.000***	-0.000	-0.000	-0.000***	-0.000
# of fam member	0.032**	0.103***	0.019	-0.058	0.029*	0.054
father's educ lvl	-0.032	-0.091	-0.081*	0.040	0.014	-0.162***
mother's educ lvl	0.038	0.077	0.070*	-0.030	0.049*	0.025
year (2003)	-0.389***	-0.208**	-0.600***	-0.344***	-0.393***	-0.360**
location (east)	0.197***	0.204**	0.131*	0.228***	0.224***	0.101

Regression results for treatment model

Table: Results of Endogenous Treatment Variable Model (y: whether parents migrate)

	All obs.	Age 0-6	Age 6-12	Age 12-18	Rural	Urban
migration rate	7.390***	6.595***	8.103***	7.503***	7.384***	7.319***
gender	-0.019	0.032	-0.022	-0.042	-0.033	0.062
age	-0.022***	0.002	-0.050**	-0.040**	-0.023***	-0.016
hhincome	-0.000***	-0.000***	-0.000	-0.000	-0.000***	-0.000
# of fam member	0.032**	0.103***	0.019	-0.058	0.029*	0.054
father's educ lvl	-0.032	-0.091	-0.081*	0.040	0.014	-0.162***
mother's educ lvl	0.038	0.077	0.070*	-0.030	0.049*	0.025
year (2003)	-0.389***	-0.208**	-0.600***	-0.344***	-0.393***	-0.360**
location (east)	0.197***	0.204**	0.131*	0.228***	0.224***	0.101

Results

- In terms of health condition of left-behind children, we found that parents' absence has negative impact to the body development, but no significant impact on "whether sick/injured".
- For education condition, children are significantly and negatively influenced as well.
- Comparing children from urban or rural area, we found that education condition is more influenced in the rural area, while health condition, height specifically is more severe among urban left-behind children.

Speaker Transition

Spatial Econometric Modeling of Origin-Destination Flows

An Analysis of 50 Million Chicago Taxi Trips

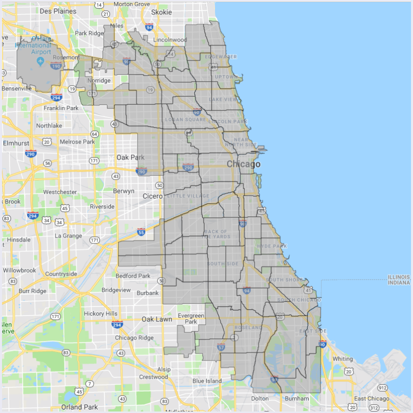
Dongping Zhang

M.A. in Computational Social Science
The University of Chicago

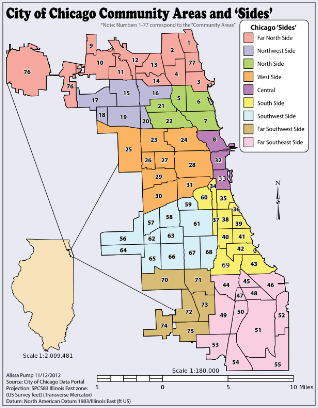
March 29, 2018

Background: Spatial Unit

Neighborhoods: Chicago 77



Community Area



Research Question

- How to model and explain variations in the level of taxi trip flows using different combinations of OD pairs?
- How to identify and measure: 1). community characteristic effects, 2). spatial separation effects and 3). temporal effects to urban interregional flows as captured by 50 million taxi trips?

Motivation

Motivation

- How people behave and interact within large and complex network system?
- How can we better explain and model collective social behaviors (mobility)?

Why care?

- To explain human mobility patterns in a highly segregated city such as Chicago.
- To identify pushing factors and pulling factors of a neighborhood.

Motivation

Motivation

- How people behave and interact within large and complex network system?
- How can we better explain and model collective social behaviors (mobility)?

Why care?

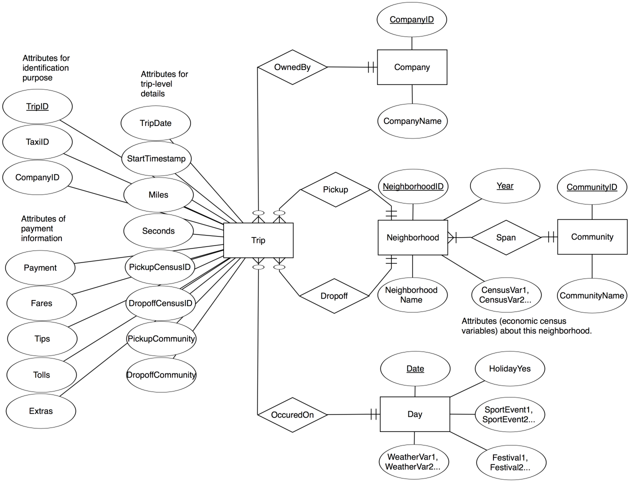
- To explain human mobility patterns in a highly segregated city such as Chicago.
- To identify pushing factors and pulling factors of a neighborhood.

Data

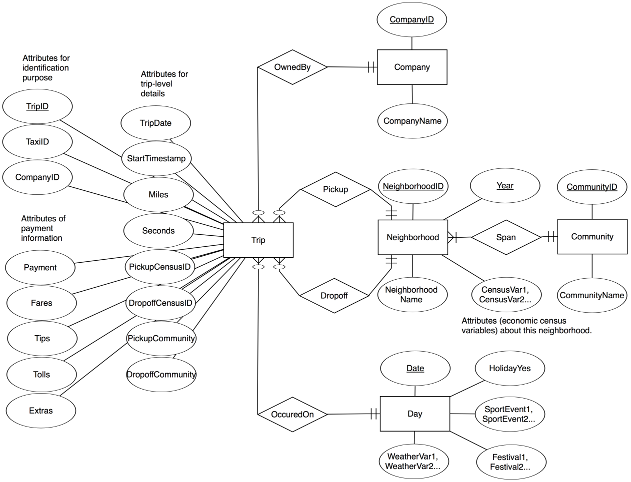
Source

- *The Chicago Data Portal*
- *The Chicago Metropolitan Agency for Planning (CMAP)*
- *National Centers for Environmental Information (NOAA)*
- *Choose Chicago* ®
- MLS, NHL, WNBA, NBA, FBS, MLB

Variables of Interest



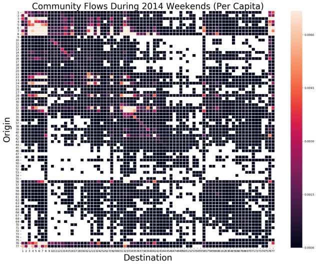
Variable Explanation



Exploratory Data Analysis: Neighborhood Flow Matrix

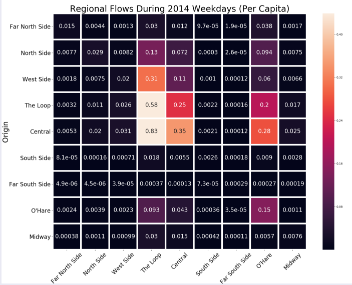
Weekday

Weekend + Holiday

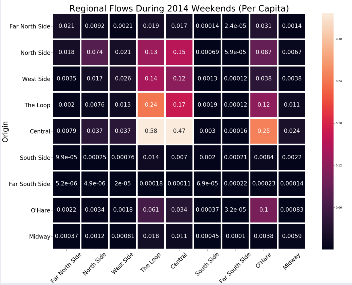


Exploratory Data Analysis: Community Flow Matrix

Weekday

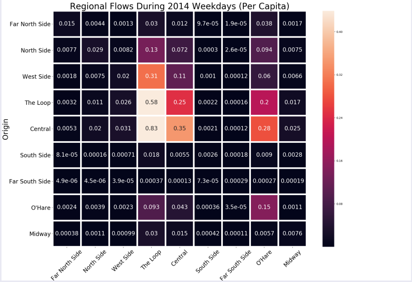


Weekend + Holiday

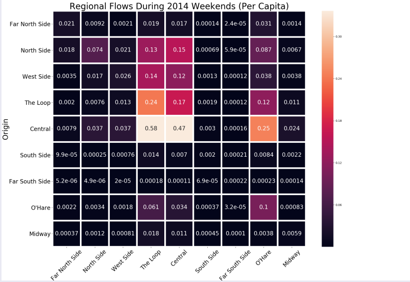


Exploratory Data Analysis: Community Flow Matrix

Weekday



Weekend + Holiday



Structural Differences of Networks

Day of a week

- Weekday
- Weekends + Holidays

Time of a day

- AM Peak (Hours: 7, 8, 9)
- Midday (Hours: 10,11,12,13,14,15)
- PM Peak (Hours: 16,17,18)
- Evening (Hours: 19, 20, 21, 22, 23)
- Early Morning (Hours: 0, 1, 2, 3, 4, 5, 6)

Structural Differences of Networks

Day of a week

- Weekday
- Weekends + Holidays

Time of a day

- AM Peak (Hours: 7, 8, 9)
- Midday (Hours: 10,11,12,13,14,15)
- PM Peak (Hours: 16,17,18)
- Evening (Hours: 19, 20, 21, 22, 23)
- Early Morning (Hours: 0, 1, 2, 3, 4, 5, 6)

How?

The Quadratic Assignment Procedure (QAP Test)

Model Specification

Classical Spatial Interaction Model of the Gravity Type

$$\mu(i,j) = CA(i)B(j)S(i,j) \tag{1}$$

$$\mu_{ij} = c \prod_{q \in Q} (A_{iq})^{\beta_q} \prod_{r \in R} (B_{jr})^{\gamma_r} \prod_{k \in K} [{}^{k}D(i,j)]^{\theta_k} \tag{2}$$

Log-normal additive

$$y(i,j) = c + \sum_{q \in Q} \beta_q a_q(i) + \sum_{r \in R} \gamma_r b_r(j) + \sum_{k \in K} \theta_k d_k(i,j) \tag{3}$$

Model Specification

Classical Spatial Interaction Model of the Gravity Type

$$\mu(i,j) = CA(i)B(j)S(i,j) \tag{4}$$

$$\mu_{ij} = c \prod_{q \in Q} (A_{iq})^{\beta_q} \prod_{r \in R} (B_{jr})^{\gamma_r} \prod_{k \in K} [{}^{k}D(i,j)]^{\theta_k} \tag{5}$$

Log-normal additive

$$y(i,j) = c + \sum_{q \in Q} \beta_q a_q(i) + \sum_{r \in R} \gamma_r b_r(j) + \sum_{k \in K} \theta_k d_k(i,j) \tag{6}$$

Matrix Notation

Compact Form

$$\boldsymbol{y} = \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{7}$$

Spatial Dependence

$$\boldsymbol{y} = \rho_o \boldsymbol{W}_o \boldsymbol{y} + \rho_d \boldsymbol{W}_d \boldsymbol{y} + \rho_w \boldsymbol{W}_w \boldsymbol{y} + \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{8}$$

Temporal Effect

$$\boldsymbol{y} = \rho_o \boldsymbol{W}_o \boldsymbol{y} + \rho_d \boldsymbol{W}_d \boldsymbol{y} + \rho_w \boldsymbol{W}_w \boldsymbol{y} + \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \phi \boldsymbol{H} + \upsilon \boldsymbol{D} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbb{I}_n) \tag{9}$$

Matrix Notation

Compact Form

$$\boldsymbol{y} = \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{10}$$

Spatial Dependence

$$\boldsymbol{y} = \rho_o \boldsymbol{W}_o \boldsymbol{y} + \rho_d \boldsymbol{W}_d \boldsymbol{y} + \rho_w \boldsymbol{W}_w \boldsymbol{y} + \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{11}$$

Temporal Effect

$$\boldsymbol{y} = \rho_o \boldsymbol{W}_o \boldsymbol{y} + \rho_d \boldsymbol{W}_d \boldsymbol{y} + \rho_w \boldsymbol{W}_w \boldsymbol{y} + \alpha \boldsymbol{\iota}_n + X_o \boldsymbol{\beta} + X_d \boldsymbol{\gamma} + D_t \boldsymbol{\theta} + \phi \boldsymbol{H} + \upsilon \boldsymbol{D} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbb{I}_n) \tag{12}$$

Distance Deterrence

Distance Deterrence Function

$$S_{(i,j)} = \prod_{k \in K} [{}^{k}D(i,j)]^{\theta_k}$$

Candidate Variables

- Euclidean Distance
- Averaged Transportation Cost/Time
- Perceived Transportation Cost/Time
- ...

Multicollinearity

Distance Deterrence Function

$$S_{(i,j)} = \prod_{k \in K} [{}^{k}D(i,j)]^{\theta_k}$$

Candidate Variables

- Euclidean Distance
- Averaged Transportation Cost/Time
- Perceived Transportation Cost/Time
- ...

Solution to Multicollinearity

Principle Component Analysis (PCA)

Parameter Estimation

Small sizes of flow

- Neighborhood-Level $= 77 \times 77 = 5929$
- Community-Level $= 9 \times 9 = 81$

Traditional Method

- Maximum Likelihood (MLE)
- Generalized Method of Moments (GMM)
- Bayesian Markov Chain Monte Carlo (MCMC)

Speaker Transition

Understanding Political Implications of Moha

Jingyuan Zhou
University of Chicago

Moha - Toad Worship


Tired
Normal
Disappointed
Happy
Thinking
Angry
Sad


Excited

Why do Chinese netizens decide to continue Moha even after many years of the berating event? What are the political implications behind this fever and what could this fever insinuate about Chinese public and their relationship with the government?

Why do Chinese netizens decide to continue Moha even after many years of the berating event? What are the political implications behind this fever and what could this fever insinuate about Chinese public and their relationship with the government?

Source materials - ‘Three Pieces About Toad’

1. Berating Hong Kong journalist event on 2000/10/27
2. ‘60 minutes’ interview with American Journalist Mike Wallace on 2000/8/15
3. Inspection at China United Engineering Corporation (Jiang worked there for some years) on 2009/4/23

Source materials - ‘Three Pieces About Toad’

1. Berating Hong Kong journalist event on 2000/10/27
2. ‘60 minutes’ interview with American Journalist Mike Wallace on 2000/8/15
3. Inspection at China United Engineering Corporation (Jiang worked there for some years) on 2009/4/23

Argument 1: Voice dissident against censorship

-Evidence: Two of three interviews happened in 2000, and people had access to the materials, but it did not become a fever until 2014 when Xi was in power.
-> comparison of censorship/stability regimes of their leadership

- Discussion: could be depoliticization - typically not understood as a challenge to Party policies

implication: govt-citizen relationship

Argument 1: Voice dissident against censorship

-Evidence: Two of three interviews happened in 2000, and people had access to the materials, but it did not become a fever until 2014 when Xi was in power.
-> comparison of censorship/stability regimes of their leadership

- Discussion: could be depoliticization - typically not understood as a challenge to Party policies

implication: govt-citizen relationship

Argument 1: Voice dissident against censorship

-Evidence: Two of three interviews happened in 2000, and people had access to the materials, but it did not become a fever until 2014 when Xi was in power.
-> comparison of censorship/stability regimes of their leadership

- Discussion: could be depoliticization - typically not understood as a challenge to Party policies

implication: govt-citizen relationship

Argument 1: Voice dissident against censorship

-Evidence: Two of three interviews happened in 2000, and people had access to the materials, but it did not become a fever until 2014 when Xi was in power.
-> comparison of censorship/stability regimes of their leadership

- Discussion: could be depoliticization - typically not understood as a challenge to Party policies

implication: govt-citizen relationship

Argument 2: Preferring Jiang's personal image -> cultural development

1. Unlike his successors' insensible or solemn political figures, Jiang's humor, interest in music, literature, language and somewhat bold behaviors in formal occasions make him **cosmopolitan, more humane and likable**. Hasi appreciate these qualities of his.

2. People miss Jiang's time when **cultural development** was more encouraged and there was less censorship in general.

Implication: values that people look up to under globalization

Argument 2: Preferring Jiang’s personal image -> cultural development

1. Unlike his successors’ insensible or solemn political figures, Jiang’s humor, interest in music, literature, language and somewhat bold behaviors in formal occasions make him **cosmopolitan, more humane and likable**. Hasi appreciate these qualities of his.

2. People miss Jiang’s time when **cultural development** was more encouraged and there was less censorship in general.

Implication: values that people look up to under globalization

Argument 3: Jiang being the figure of highest power and the person who challenges presumed rules -> sense of ambiguity

Hasi agree with viewing the world as more complicated than black-and-white but enjoy the grey area in between.

implication: tired of black-white/extremist political environment and prefers the sense of ambiguity

Argument 3: Jiang being the figure of highest power and the person who challenges presumed rules -> sense of ambiguity

Hasi agree with viewing the world as more complicated than black-and-white but enjoy the grey area in between.

implication: tired of black-white/extremist political environment and prefers the sense of ambiguity

Hasi are dissident to elevated censorship and especially its control over cultural development; however, there is radical undecidability of how political these memes are and will become. It could also suggest that Chinese citizens are dissatisfied with the current government, but due to the power the government is able to demonstrate, they are afraid of taking their dissidence into action and stand up against it.

Hasi are dissident to elevated censorship and especially its control over cultural development; however, there is radical undecidability of how political these memes are and will become. It could also suggest that Chinese citizens are dissatisfied with the current government, but due to the power the government is able to demonstrate, they are afraid of taking their dissidence into action and stand up against it.

Why ambiguity has become the essence of Moha culture

Hasi need to express

- their ideas against censorship
- their cosmopolitan values
- their view of interpreting the world as more complicated than black-and-white

but they also need to protect their posts from being censored and themselves for being invited to talk with the officials (or to check the water meter).

Hasi need to express

- their ideas against censorship
- their cosmopolitan values
- their view of interpreting the world as more complicated than black-and-white

but they also need to protect their posts from being censored and themselves for being invited to talk with the officials (or to check the water meter).

Speaker Transition

NEGOTIATION ON HEGEMONIC NOTIONS OF GENDER IN AMERICAN CINEMA AFTER 2008 GREAT RECESSION

YIQING ZHU

IMPORTANCE OF UNDERSTANDING GENDER ROLES

- "We the People."
- White patriarchal capitalism
- Gender roles and expectations permeate our culture, language, and media in ways both subtle and obvious.

IMPORTANCE OF UNDERSTANDING GENDER ROLES

- "We the People."
- White patriarchal capitalism
- Gender roles and expectations permeate our culture, language, and media in ways both subtle and obvious.

IMPORTANCE OF UNDERSTANDING GENDER ROLES

- "We the People."
- White patriarchal capitalism
- Gender roles and expectations permeate our culture, language, and media in ways both subtle and obvious.

IMPORTANCE OF STUDYING POPULAR CULTURE

- We get ideas about what it means to be a boy or a girl from ideological institutions such as the family, the schools, other children, and the media.
- Many theorists feel that in today's electronic world, the media has more influence on cultural ideas and ideologies than do schools, religions, and families combined.

IMPORTANCE OF STUDYING POPULAR CULTURE

- We get ideas about what it means to be a boy or a girl from ideological institutions such as the family, the schools, other children, and the media.
- Many theorists feel that in today's electronic world, the media has more influence on cultural ideas and ideologies than do schools, religions, and families combined.

WHY HOLLYWOOD FILMS?

- Hollywood film is so prevalent in American culture.
- Hollywood promotes its films not as political tracts but as mindless escapism, and an audience member who accepts that tenet will rarely be alert to the cultural and ideological assumptions that the films encode and promote.

WHY HOLLYWOOD FILMS?

- Hollywood film is so prevalent in American culture.
- Hollywood promotes its films not as political tracts but as mindless escapism, and an audience member who accepts that tenet will rarely be alert to the cultural and ideological assumptions that the films encode and promote.

HEGEMONIC PATRIARCHY IN AMERICAN FILMS

- Contemporary Hollywood filmmaking, by drawing on certain representational patterns and formulas left over from previous decades, continues to marginalize women and women’s issues while both subtly and forthrightly privileging men and masculinity.

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

- Social events are catalysis of social construction of gender.

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

During the Great Depression:

- Strong female roles are curtailed and men are rougher and tougher
- Woman's films
- screwball comedy – push the reversal of gender roles

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

During the World War II:

- Strong working woman: Rosie the Riveter
- Reinforce the triumphant masculinity
- Promote team work spirit
- Suppress masculine individuality

Rosie the Riveter

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

After the World War II:

- Shift woman back home
- film noir & blonde bombshell – sexualized woman
- social problem films – displaying cracks in masculine confidence

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

After the World War II:

- Shift woman back home
- film noir & blonde bombshell – sexualized woman
- social problem films – displaying cracks in masculine confidence

HEGEMONIC NEGOTIATION IN AMERICAN FILMS

- Social events are catalysis of social construction of gender.
- **Since the Great Recession is the worst global recession since the Great Depression, how the films during and after the Great Recession period reflect the changing socio-cultural construction of gender?**

DATA

IMDB:

- 31,351 US feature films released in 2007 - 2017
- **Text data**: synopsis, plot summary, keyword
- **Attribute data**: release year, genre, cast, domestic box office receipt

DATA

Synopsis

In Santa Cecilia, Mexico, Imelda Rivera was the wife of a musician who left her and their 3-year-old daughter Coco, to pursue a career in music. She banned music in the family and opened a shoe-making family business. Ninety-six years later, her great-great-grandson, 12-year-old Miguel, now lives with Coco and their family. He secretly dreams of becoming a musician like Ernesto de la Cruz, a popular actor and singer of Coco's generation. One day, Miguel inadvertently damages the photo of Coco with her parents at the center of the family ofrenda and removes it, discovering that her father (whose face is torn out) was holding Ernesto's famous guitar.

Concluding that Ernesto is his great-great-grandfather, Miguel ignores his grandmother Elena's objections and leaves to enter a talent show for the Day of the Dead. He enters Ernesto's mausoleum and steals his guitar to use in the show, but becomes invisible to everyone in the village plaza. However, he can see and be seen by his Xoloitzcuintli dog Dante and his skeletal dead relatives who are visiting from the Land of the Dead for the holiday. Taking him there, they realize that Imelda cannot visit as Miguel removed her photo from the ofrenda. Discovering that he is cursed for stealing from the dead, Miguel must return to the Land of the Living before sunrise or he will become one of the dead: to do so, he must receive a blessing from a member of his family using an Aztec marigold petal that can undo the curse placed upon him by stealing Ernesto's guitar. Imelda offers Miguel a blessing but on the condition that he abandon his musical pursuits when he returns to the Land of the Living; Miguel refuses and attempts to seek Ernesto's blessing.

Miguel encounters Héctor, a down-on-his-luck skeleton who once played with Ernesto and offers to help Miguel reach him. In return, Héctor asks Miguel to take his photo back to the Land of the Living so he can visit his daughter before she forgets him and he disappears completely. Héctor attempts to return Miguel to his relatives, but Miguel escapes and infiltrates Ernesto's mansion, learning along the way that an old friendship between the two deteriorated before Héctor's death. Ernesto welcomes Miguel as his descendant, but Héctor confronts them, imploring Miguel to take his photo. Miguel soon realizes that Ernesto murdered Héctor using a poisoned drink and stole the songs he had written, passing them off as his own to become famous. To maintain his legacy, Ernesto steals the photo and has Miguel and Héctor thrown into a cenote pit.

Miguel realizes that Héctor is his actual great-great-grandfather and that Coco is Héctor's daughter, the only living person who still remembers him. With the help of Dante - who turns into an alebrije - the dead Riveras find and rescue them. Miguel reveals that Héctor's decision to return home to her and Coco resulted in his death, and Imelda and Héctor reconcile. They infiltrate Ernesto's sunrise concert to retrieve Héctor's photo from Ernesto and expose his crimes. Ernesto is crushed by a falling church bell as in his previous life, but the photo falls into the water and disappears.

Synopsis

Plot Summary

The rebellious Miguel, a 12-year-old Mexican boy and hopeful musician, is unable to understand the family's continuing ban on all music, especially when his icon and the greatest guitar player to ever live, the deceased Ernesto de la Cruz, is the town's hero. However, an inadvertent mistake on the sacred Day of the Dead will magically transport Miguel to the distant and bustling Land of the Dead, where the scoundrel skeleton, Hector, will lead the way through the vibrant underworld to help the young trespasser find a missing ancestor. But can they do it before sunrise?

—*Nick Riganas*

Keyword

mexico 3 of 3 found this relevant	ban on music 1 of 1 found this relevant
12 year old 1 of 1 found this relevant	afterlife 1 of 1 found this relevant
ghost 1 of 1 found this relevant	male protagonist 1 of 1 found this relevant
day of the dead 1 of 1 found this relevant	cgi animation 1 of 1 found this relevant
disney 1 of 1 found this relevant	one word title Is this relevant?
voice over Is this relevant?	voice over narration Is this relevant?
family relationships Is this relevant?	musician Is this relevant?
shoemaking Is this relevant?	young boy Is this relevant?
boy Is this relevant?	great grandmother Is this relevant?
ancestor Is this relevant?	bell Is this relevant?
bell falling on someone Is this relevant?	dia de los muertos Is this relevant?

DATA

Why synopsis instead of script?

- User collaborated synopsis is the widely accepted and important information of encoded intention from films.
- Not "reading too much into things".

METHOD

Computational Content Analysis:

- Corpus Linguistics
- Clustering and Topic Modeling
- Word Embedding
- Information Extraction
- Semantic Networks

MEASUREMENT

Compare by release year, genre, leading role gender, woman's / man's movies:

- Screen time: frequency of characters of different gender
- Representation: grams / name-entities associated with woman and man,

  semantic distance between certain words and woman / man

  document similarity