



# Open Source Toolkit Development for Spatial Social Network Analysis and Simulation

Xinyue Ye, Jay Lee, and Sagar Naresh Jha

August 12, 2015



**Spatiotemporal Modeling of Human Dynamics Across Social Media and Social Networks  
Interdisciplinary Behavioral and Social Science Research, National Science Foundation**



# Theoretical Motivation

- Integration of space and time generates much closer interactions among social sciences (Goodchild et al., 2000; Goodchild, 2006).
- Treatment of spatial and temporal factors will determine how the reality and the theoretical framework are examined (Ye & Carroll, 2011; Ye & Rey, 2013).



# Methodological Motivation

- Imperative need for effective and efficient methods to represent and examine the coupled space-time attributes of socioeconomic and human dynamics phenomena in the comparative context (Janikas, 2007; Rey & Ye 2010; Ye & Rey, 2013; Ye et al., 2014).



## **Anselin (2011): three important aspects that stimulated spatial analytical software development in the 21st century**

- the role of **methodological** innovations in both exploratory and confirmatory analysis.
- the role of the **open source** software movement in stimulating new development and broadening the community of developers and adopters.
- the role of the **internet**, in the form of web-based spatial analysis, spatial analytical web services, and the advent of a scientific cyberinfrastructure for geospatial analysis.

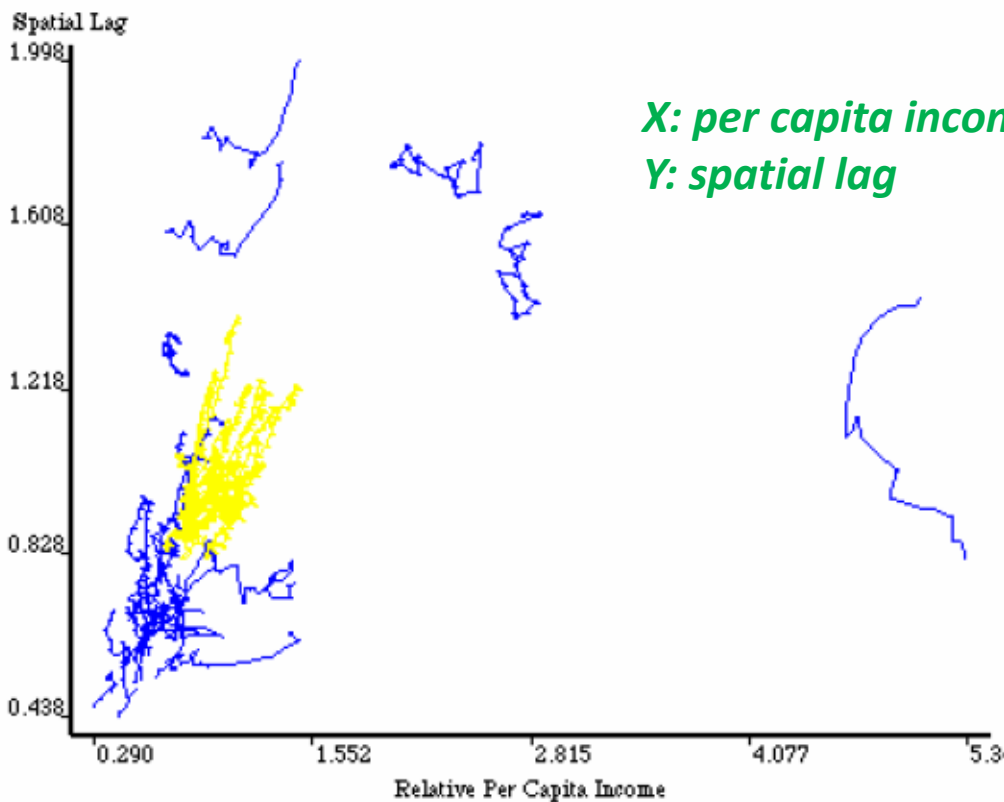


# User-led Innovation

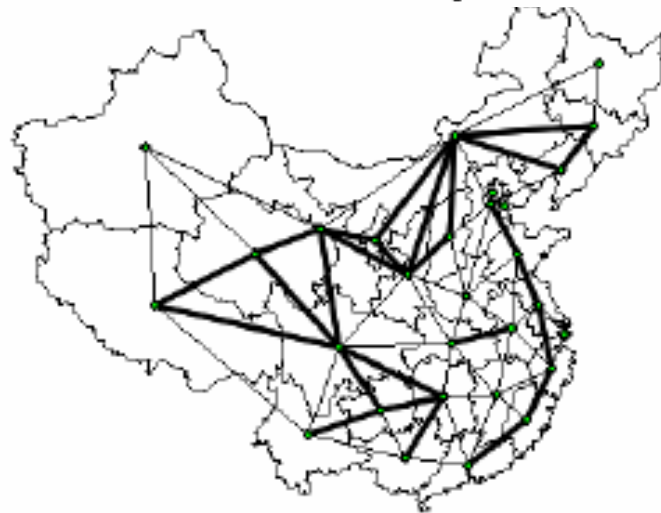
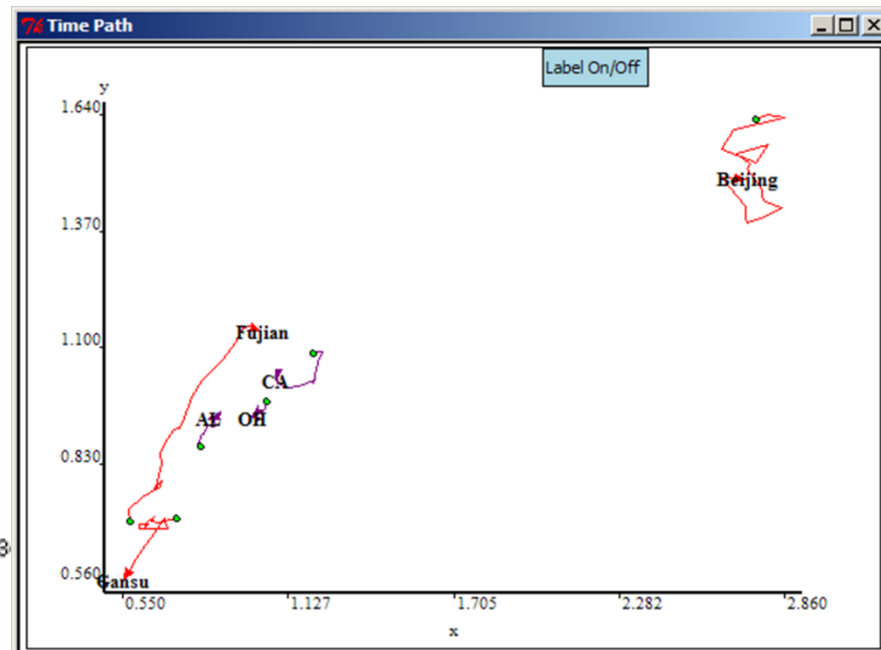
- Open source represents a paradigm shift in geospatial research that has facilitated collaboration across disciplines.
- The development of open source GIS packages has been boosted. However, open source projects in the areas of advanced spatial analysis are very few (Rey & Anselin, 2007; Rey, 2009; Ye & Rey, 2013).

# STARS: Space Time Analysis of Regional Systems

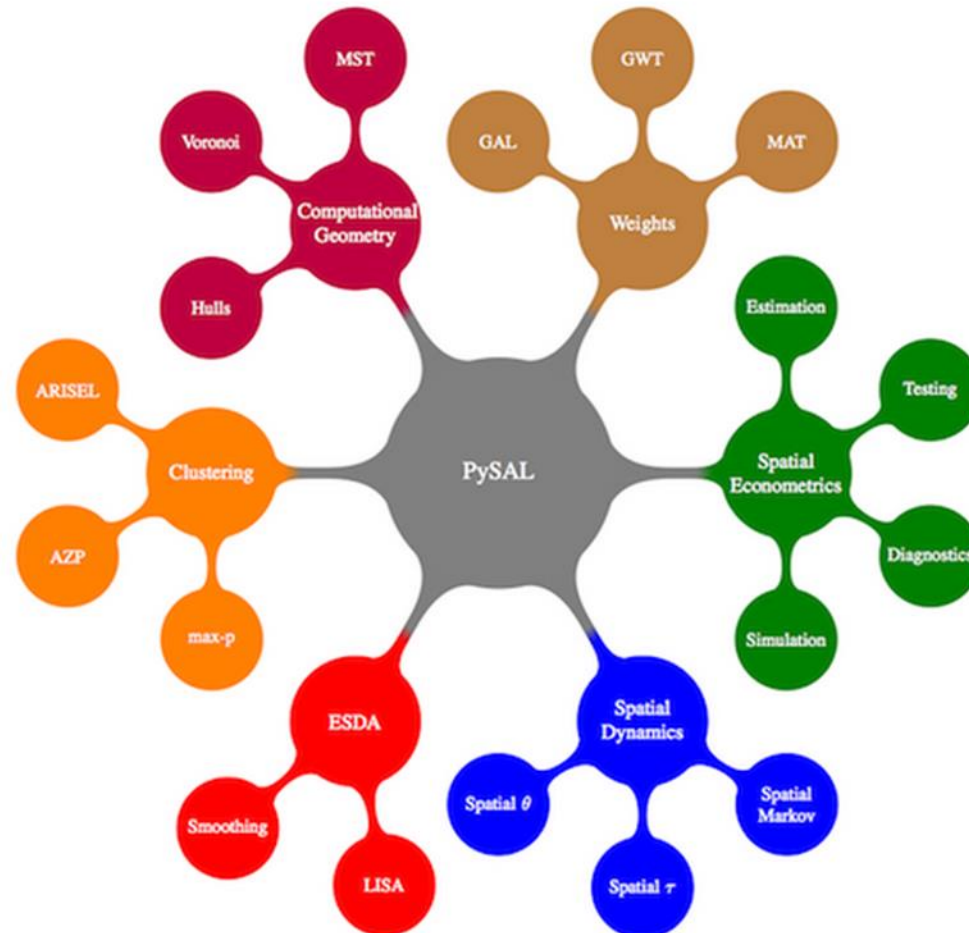




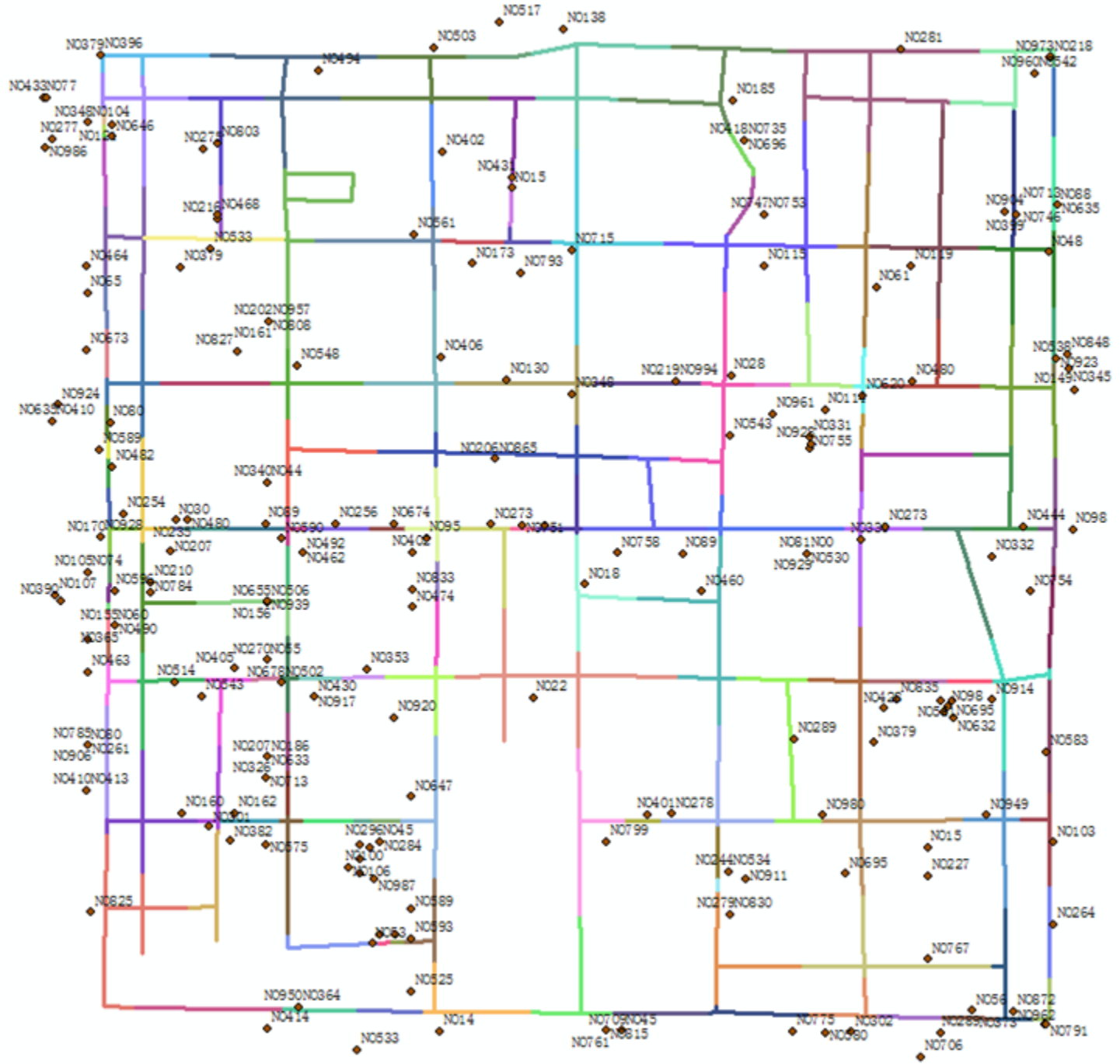
*X: per capita income relative to national mean*  
*Y: spatial lag*



# PySAL: Open Source Python Library for Spatial Analytical Functions









### 74 Ohio Warn Notice Analysis

Frame  
 Canvas Width: 600 pixels  
 Canvas Height: 600 pixels  
 Set Default

File Pre-processing  
 Convert PDF to CSV

File Operation  
 Load a GIS Polygon Shape File  
 Load a GIS Point Shape File  
 Load a GIS Polyline Shape File  
 Load a CSV Data File  
 Load a PDF Data File

### Logical Error checking -- Ohio Warn Notice

You can edit relevant records in CSV file, and load CSV file next time.

Data Rcvd Error:  
 ==>3/26/2009<== ABX Air, Inc. Wilmington (Clinton) 55 page 9

City(County) Error:  
 11/12/2009 Guardian Automotive Products, Inc. ==>Upper Sandusky<== 59 page 2  
 4/27/2009 Harco Manufacturing Group, LLC ==>Moraine<== 106 page 8

Lay-off Number Error:  
 7/30/2009 NCR Corporation Dayton (Montgomery) ==>644 \*updated<== page 4  
 7/17/2009 Severstal Warren, Inc. Warren (Trumbull) ==>1,263<== page 4  
 7/15/2009 ABX Air, Inc. Wilmington (Clinton) ==>1,034<== page 4  
 6/30/2009 BBI Enterprises Group, Inc. Sidney (Shelby) ==>?<== page 5  
 6/5/2009 Apria Healthcare Minster (Auglaize) ==>Unknown<== page 6

Open PDF File to Check    Open CSV File to Edit Error

Convert

Successfully Convert to CSV

OK



### 74 Ohio Warn Notice Analysis - Geographic Information System Image

Frame  
 Canvas Width: 600 pixels  
 Canvas Height: 600 pixels  
 Set Default

File Pre-processing  
 Convert PDF to CSV

File Operation  
 Load a GIS Polygon Shape File  
 Load a GIS Point Shape File  
 Load a GIS Polyline Shape File  
 Load a CSV Data File  
 Load a PDF Data File

Canvas Operation  
 Zoom in    Reset  
 Clear Legend

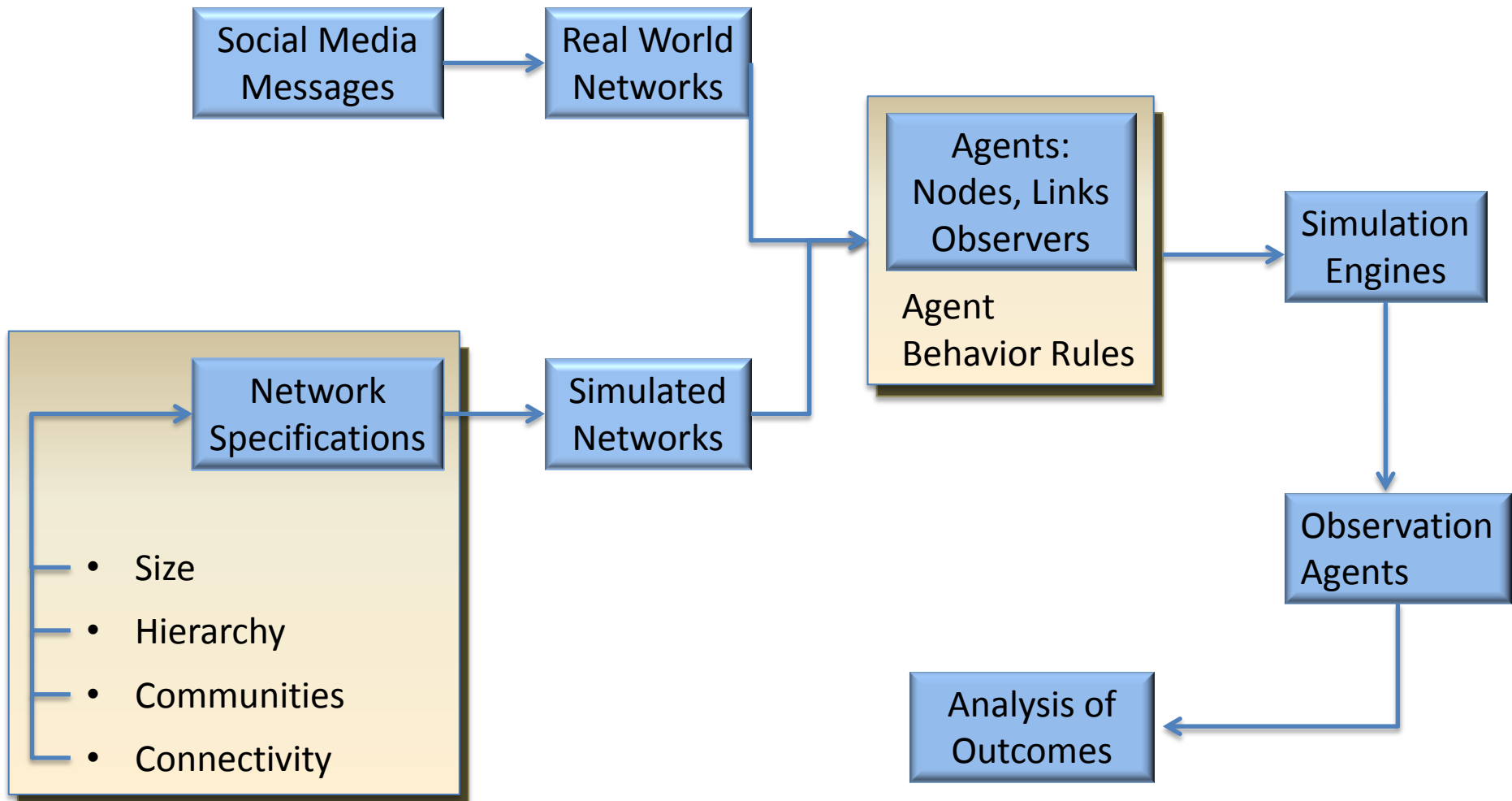
Location  
 Akron (Summit)  
 Alliance (Stark)  
 Archbold (Fulton)  
 Aurora (Portage)  
 Austintown (Mahoning)  
 Barberton (Summit)  
 Bedford (Cuyahoga)  
 Bellefontaine (Logan)  
 Bowling Green (Wood)  
 Brilliant (Jefferson)

Show    Undo  
 Select    Detail

# Spatiotemporal Analysis and ABM

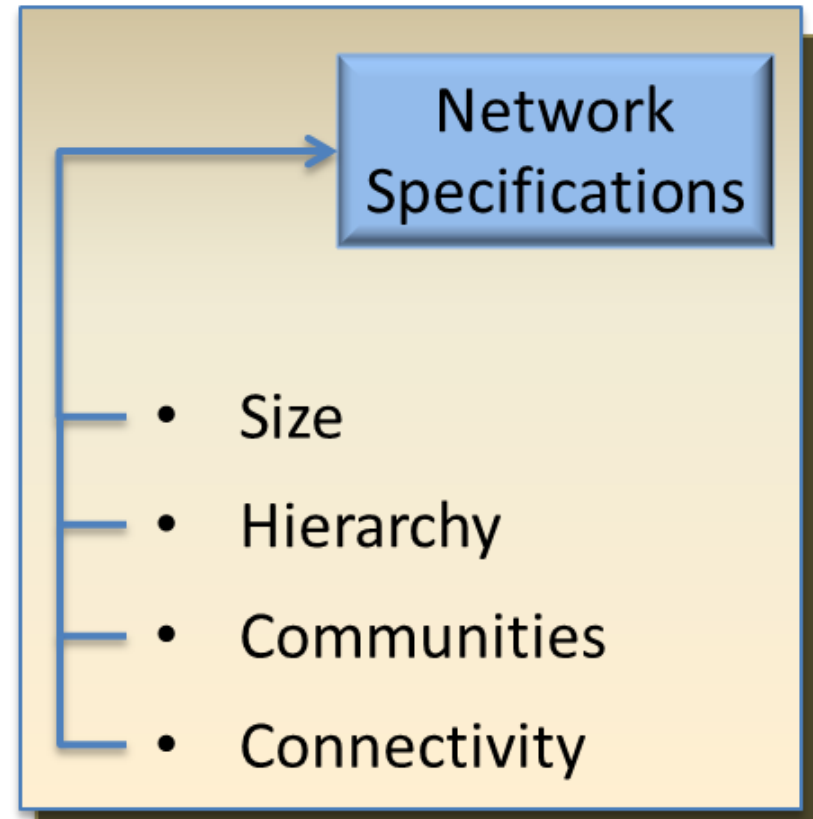
- ❖ **Agent-based model (ABM)** is one of a class of *computational models* for *simulating* the actions and **interactions** of autonomous agents (both individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole. (*Wiki*)
- ❖ Each node and link in the networks are individual agents.
- ❖ The overall networks is the system as a whole.
- ❖ Diffusion is the **interactions** in the **space-time-network** context.
- ❖ Rule development and result verification using **space-time analysis**.

# Overall Research Design of Space-Time-Network Tool



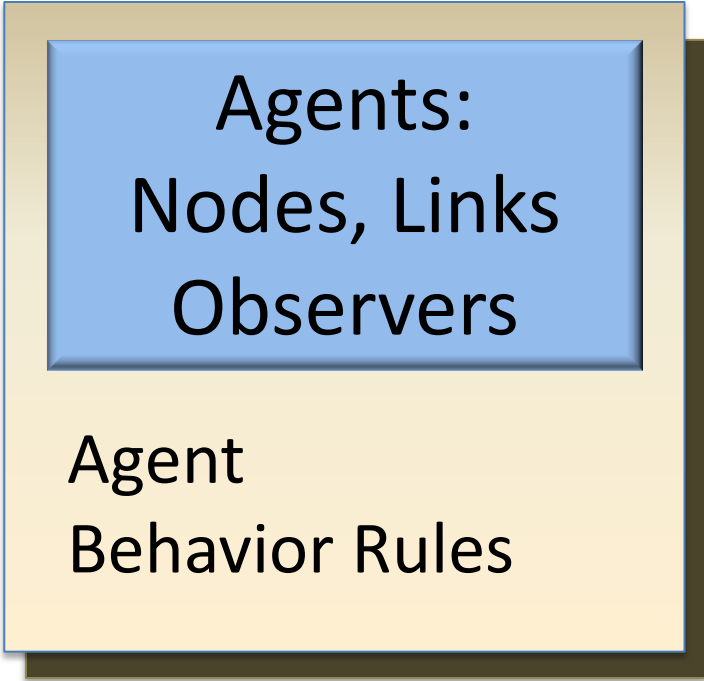
# Network Specifications

- **Size**
  - # nodes
  - # links
- **Hierarchy**
  - # tiers
  - # level 1 opinion leaders
  - # level 2 opinion leaders
  - .....
- **Communities**
  - # communities
- **Connectivity**
  - Connectivity index



# Model Agents

- **Agents**
  - Nodes
    - e.g., Twitter accounts
    - e.g., Opinion leaders
    - e.g., Community interfaces
  - Links
    - e.g., Re-tweet
    - e.g., Passage/Blockage
  - Observer
    - e.g. Statistical summaries



Agents:  
Nodes, Links  
Observers

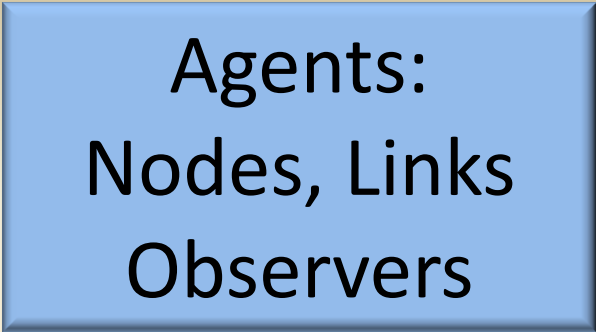
Agent  
Behavior Rules

# Agent Rules

- **Behavioral rules**

- Tier 1/Tier 2/Tier 3 opinion leaders
  - e.g., Probabilities for retweeting
  - e.g., Received messages
- Links
  - e.g., Passage
  - e.g., Blockage
- Communities
  - e.g., Connectivity

Simulate and monitor the step-by-step process that a meme would flow from “opinion leaders” (the hubs of the network) to “followers” (those connected to the hubs directly or indirectly) in the network.



Agents:  
Nodes, Links  
Observers



Agent  
Behavior Rules

# Tool 0.1



NLTK 3.0 documentation

[NEXT](#) | [MODULES](#) | [INDEX](#)

## Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](#) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, and an active [discussion forum](#).



## NetworkX

[NetworkX Home](#) | [Documentation](#) | [Download](#) | [Developer \(Github\)](#)

High-productivity software for complex networks

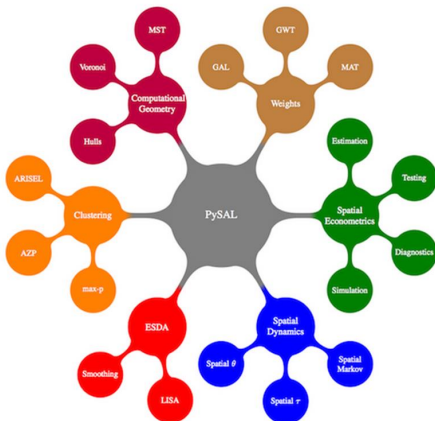
NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



[Documentation](#)  
*all documentation*

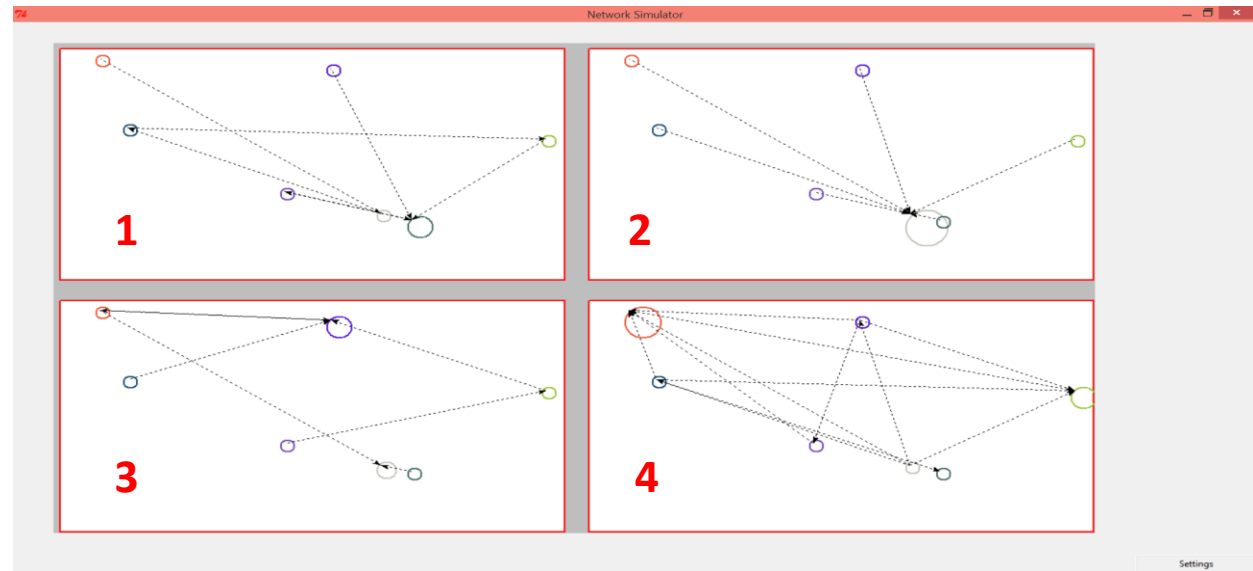
[Examples](#)  
*using the library*

[Reference](#)  
*all functions and methods*





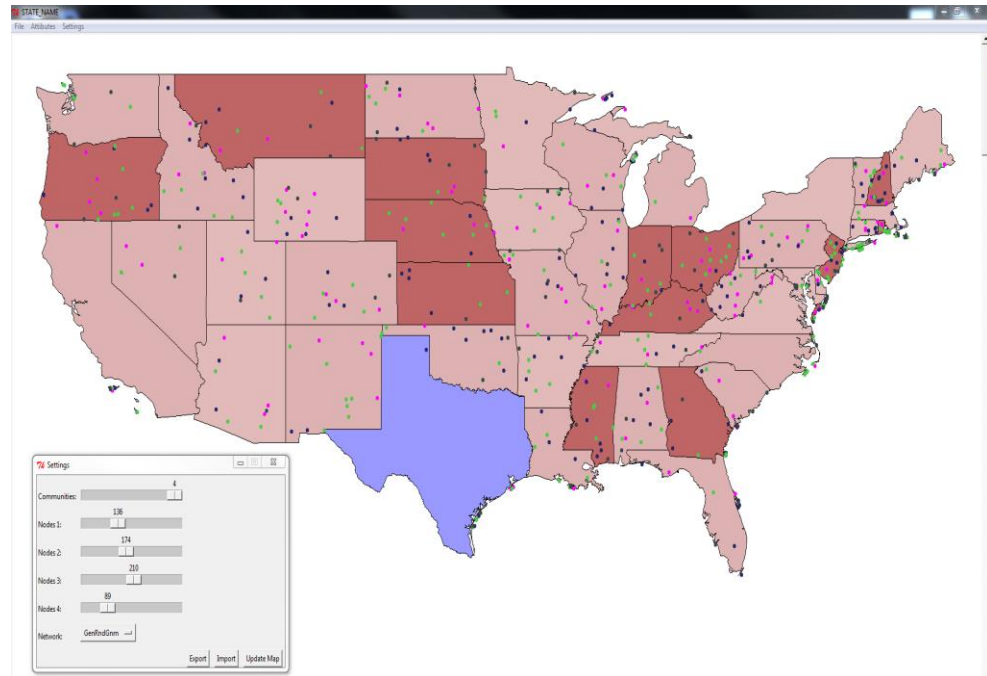
# Network Simulator – V1



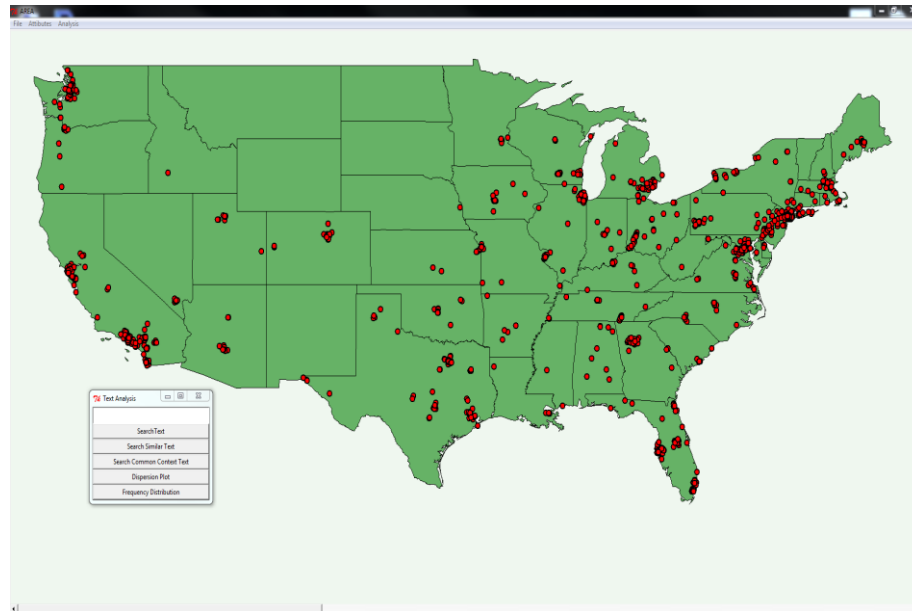
- Four examples for different types of networks
  - 1 : Random network
  - 2 : Star topology-based network
  - 3 : Endos – Renyi random network
  - 4 : Directed graph network
  - *Combinations?*

# Network Simulator – V2

- Implemented so far
  - Import and visualize shape files (.shp file) on canvas.
  - Simulate networks on user defined parameters.
  - Import and visualize real world twitter data.
  - Perform text analysis on the twitter data.



# Plot real world twitter data



_keyword	platform	geo_lat	geo_lon	place	country	tweet_in	retweet_c	favourite	user_id	user_scre	user_nam	user_locat	user_geoc	user_follc	user_frier	text
4sq	<a href="f	43.03823	-87.9434	Milwaukee	United States	16726631	0	1	2.21E+08	Schroeder	John Schroeder	TRUE	537	642	@fm1021:	
4sq	<a href="f	34.04836	-84.5988	Kennesaw	United States		0	0	13871082	wfowlkes	Will Fowl	Kennesaw	TRUE	1537	1999	Working o
4sq	<a href="f	42.51984	-92.4555	Cedar Fall	United States		0	0	14852278	Naanad	Naanad	Cedar Fall	TRUE	560	730	Time for s
4sq	<a href="f	28.0043	-82.4482	Tampa, FL	United States		0	0	15818129	GaytorKer	Ken Key	ÃœT: 27.9	TRUE	476	884	I'm at Ken
4sq	<a href="f	43.08281	-88.3628	Hartland, WI	United States		0	0	15674417	charlesrat	charlesrat	Hartland, WI	TRUE	552	1006	I'm at Hart
4sq	<a href="f	40.26327	-80.1315	Thompson	United States		0	0	2.51E+09	kopitarus	Shelby	The 412 vi	TRUE	158	463	I'm at @Pz
4sq	<a href="f	41.91037	-87.6881	Chicago, IL	United States		0	0	26016096	Chandlers	Chandler	Logan Squ	TRUE	296	406	I'm at @Rz
4sq	<a href="f	33.58447	-111.923	Scottsdale	United States		0	0	15285490	OliviaWilc	Olivia Wil	AZ and CA	TRUE	7820	2209	I'm at Hark
4sq	<a href="f	44.90976	-89.5969	Weston, WI	United States		0	0	1.32E+09	cl_hutton	Carrie Hut	Wausau	TRUE	342	829	Monthly G

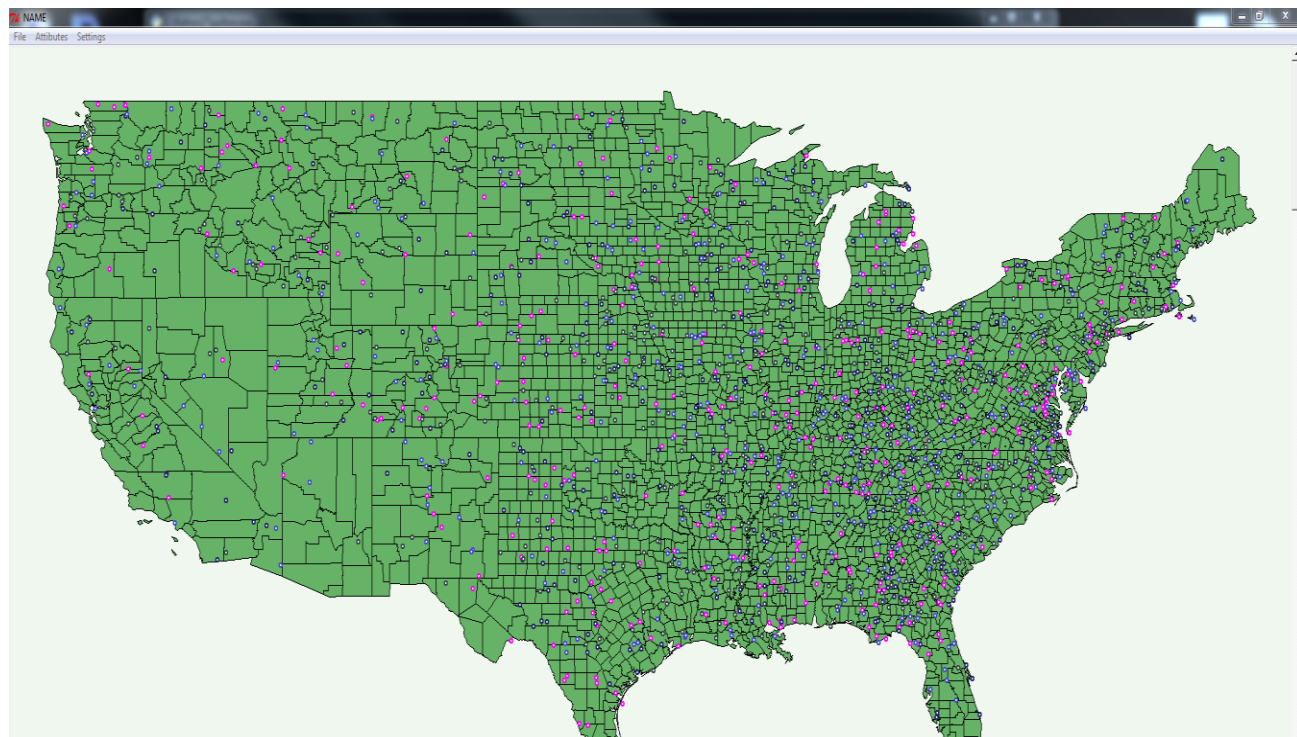
## Classical networks

- Implemented classical networks (**SNAP Networks**)
  - GenRndGnm:
    - Generates an Erdos-Renyi random graph of directed type.
  - GenForestFire:
    - Generates a random Forest Fire, directed graph with given probabilities.
  - GenStar:
    - Generates a graph with star topology. It will have a node connected to all other nodes of the network.
- More implementations are planned

## Graph Generators

- GenFull
- GenCircle
- GenGrid
- GenStar
- GenTree
- GenRndGnm
- GenPrefAttach
- GenGeoPrefAttach
- GenForestFire
- GenSmallWorld
- GenBaraHierar
- GenConfModel
- GenConfModel
- GenCopyModel
- GenDegSeq
- GenRewire
- GenRndDegK
- GenRndPowerLaw
- GenRMat
- GenRMatEpinions

# Simulated Network



Nodes only

avoids crowded display

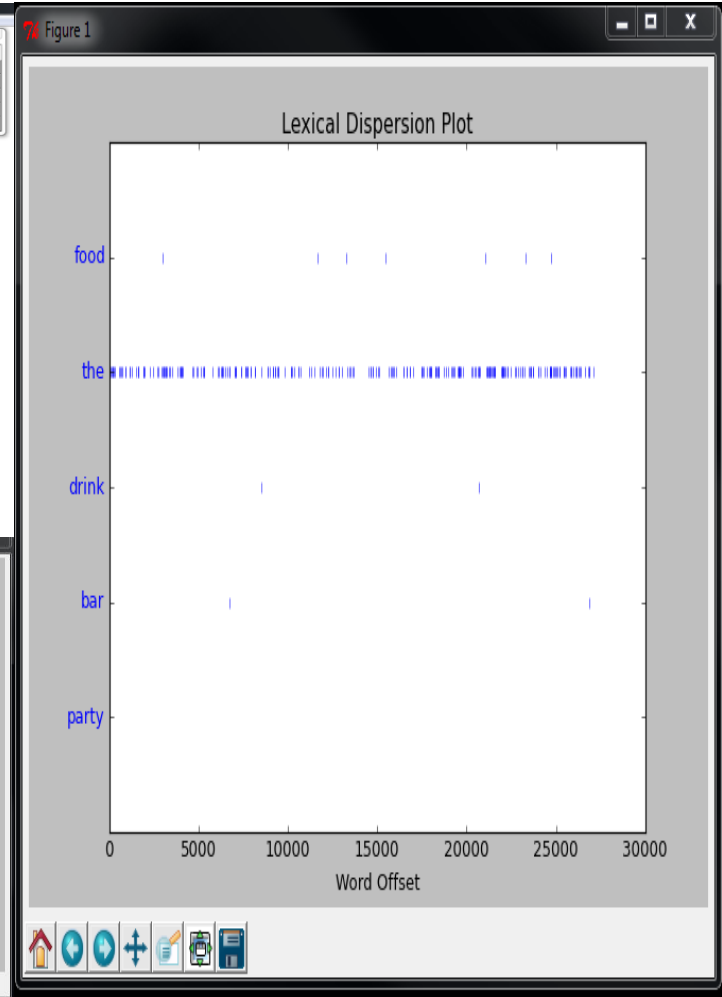
Colors represent communities

connectivity depends on connections between communities

# Text Analysis

```
Python 2.7.10 (default, May 23 2015, 09:44:00) [MSC v.1500 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
>>>
>>> open shape file!
C:/Users/sjhal/Documents/GitHub/Network_Analysis/Tool_US_Map/Tool/ShapeFiles/states/tweetdata_1.shp
fileLength, number of units: 9130 2270
Search Text
-----
Displaying 16 of 16 matches:
da24U At Sylvia's, the queen of soul food tonight! Corn bread, Cat fish and fr
n Buffalo, NY http://t.co/9W0Qc0lm3s Food Truck Rodeo!!!!!! (@ Southern Twer
tp://t.co/7PUJL1XztG I'm at Jimmy's Food Store in Dallas, TX http://t.co/jzu6
; Family! #Vindaloo (@ Mantra Indian Food in Temecula, CA) http://t.co/UOQk7WY
place brings back memories. awesome food and good times! (@ Blue Moon Cafe in
Night tonight - most amazing Mexican food I've had! @rockinghorsenyc down in..
Love Nino's Mid Town location! Great food and I love their olives. (at @NinosP
sy, TN http://t.co/Yo6n133AKi I'm at Food Court - @smallincolumbia in Columbia,
p://t.co/vpEQz22h3 I'm at Woodman's Food Market - @woodmansFood33 in Carpente
p://t.co/98380JEB3f I'm at Woodman's Food Market - @woodmansFood16 in Madison,
VMYJ78J0 Hanging out w/ my bud. Good food & drinks = good times! (@ Cork F
d & drinks = good times! (@ Cork Food & Drink in Tyler, TX) http://t.c
p://t.co/coUrW2WJ08 I'm at Breakfast food cart (New York, NY) http://t.co/Daaf
//t.co/EZMq0oR36 Out for the second food run of the month. (@ WINGO Foods in
KMDtaQW6TY I'm at Bridgeport Pasty - Food Truck in Chicago, IL http://t.co/9Po
rport) http://t.co/axK9c5eGUC I'm at Food Truck Invasion's Family Night @ Plan
-----
Search Similar Text
-----
vans neighborhood supreme general
-----
Search common context
-----
at_store at_truck breakfast_cart great_and pasty_truck second_run
ny_truck mexican down_at_market good_drinks_at_court awesome_and
indian_in_cork_drink soul_corn
-----
Search common context
-----
No common contexts were found
```

The plot shows cumulative counts on the y-axis (ranging from 2000 to 10000) and an unlabeled x-axis (ranging from 0 to 30000). The curve starts at approximately (0, 2000) and rises steeply, then levels off as it approaches 10000.



# Demo (Sandbox)

