A Lotka-Volterra Competition Model of Social Media Sites

By: Aditya Patel

Courant Institute of Mathematics

DCSM Workshop

May 15, 2012

Goal

To model competition between social media sites, in terms of users, by using the Lotka-Volterra equations

Outline

- Goal
- Introduction
  - Background
  - Question
- The Model
  - Parameters
  - Equations
  - Data
  - Analysis
- Results
  - Fixed Parameters
  - Stochastic Process
- Substantive Results
- Moving Forward

Background

- Across all social media sites, there are approximately a billion users.
  - 526 million daily Facebook active users in March 2012.
  - 800+ million unique users visit YouTube each month.
  - 6+ million Tumblr users.
- Social Media is used as a marketing tool to promote a party's product (consumer goods, ideas, etc) over competitors'.
- Social media has about 2.5 more long-term elasticity than traditional advertising methods. (Trusov Bucklin Pauwels 2008)
- An increase of unique users of a social media site translates to an increase in the consumption of the product. (Huang Singh Ghose 2011)

Uses for Social Media:

- Enhance Brand Awareness
- Protect Your Party's Reputation
- Strengthen Public Relations
- Extend a Party's Customer Service
- Facilitate Research and Development
- Build Community of Advocates
- Driving Sales and Leads (Winner-Take-All)

Question

How can we model the number of unique users of a Social Media Site while it is interacting with other sites?

Subsequent Questions

- What factors determine the intrinsic growth rate?
  - Technology, Innovation, Network Expansion, etc.
- How are the interactions with other social media sites determinant of the unique site visits?
  - Establishing a differentiated niche role in the market.
- How are internal interactions determinant of the user population?
  - Like-minded users tend to congregate with each other.

Definitions and Terms

- Unique Users: Measures the popularity of a website over a period of time. Usually requires an IP address and the cookies that correspond to the user activity.
- User-Generated Content: All content on a social media site is produced solely by user interactions. The site just provides the tools to allow users to communicate with each other.
- Winner-Take-All: The main idea behind a Social Media Site is to obtain the maximum number of unique users, where the maximum number is all the users of a population.

The Model

- Two Lotka-Volterra competition equations with six parameters.
- Use Lotka-Volterra because one can translate biological population models to SMS user growth models.
  - Two species competing for the same food source is the same idea as for two sites competing for the pool of user.
- Not many biological models are used to model competition between SMS.

The Model

$$\frac{dP_x}{dt} = P_x\left(a_x + b_x P_x + c_{xy} P_y\right)$$

$$\frac{dP_y}{dt} = P_y\left(a_y + b_y P_y + c_{yx} P_x\right)$$

$$c_{ij} = \begin{bmatrix} b_i & c_{ij} \\ c_{ji} & b_j \end{bmatrix}, \quad c_{ij} \neq c_{ji}$$

Here K_i is the carrying capacity for SMS *i*. $$b_i = \frac{a_i}{K_i}$$

The Model

- First Model: The coefficients are fixed, which models a specific condition given a time interval, and is based on set parameters. These coefficients are fitted for the data.
- Second Model: An unique model, which is based off the initial Lotka-Volterra model presented, but extends off the first model by incorporating a hierarchal model. The interaction between the two sites is now assumed to be a stochastic process. The external inter action rate, c_{ij}, is not fixed and continuously evolving. This coefficient was not directly determined by the data.
  - One model examines c_{ij} was a biased random walk

The Parameters

- Who: Two competing Social Media Sites (X and Y).
- What: Competing over unique users. This model assumes there are a set number of users in the system and a user can only be a user of X or Y.
- When: Over discrete time-steps (d_t), modeled over a year's span.
- Why: To maximize the number of unique users at the end of the time steps (X and Y).
- How: Various interactions, both internal and external (b_i and c_{ij}) that influence growth.

The Parameters

<u>**Exogenous Growth Rate (a_i)**</u>: Intrinsic rate of growth for the site, meaning this is the rate at which the site increases the number of unique users.

-Positive a_i: increasing rate of unique users

-Negative a_i: decreasing rate of unique users

<u>**Internal Interaction Rate (b_i)**</u>: This parameter represents the rate that user interactions within the community promote growth or decay. This rate represents the efficiency of the population based on available resources.

-Negative b_i: Loss of users based on internal interactions. Only a negative rate was observed from the data.

<u>**External Interaction Rate (c_{ij})**</u>: This is the rate that determines the loss or gain of users based on its interactions with other sites. It is important to note that the rate in which A interacts with B does not equal the rate at which B interacts with A.

-Positive c_{ij}: Rate of gain of users from Site J

-Negative c_{ij}: Rate of loss of users to Site J

Data

- Collected from www.siteanalytics.compete.com, which counts and calculates the number of Unique Users of a site within a week (discrete time steps).
  - The time frame for collection is Monday 12:00 AM to Sunday 11:59 PM.
- Tracked the number of unique visitors that logged on to eHarmony.com and Match.com since January 2011 by creating an Microsoft Excel Spreadsheet.
  - Every week, starting from Week 1, corresponded to an exact number of UU that site received from the previous week.
- These sites had the most data available, hence are often sited in literature, like Zhang Sarvary 2011.

Data

- The data is representative of the change in the amount of UU over a discrete span of time.
  - This showed the growth trends of eHarmony.com and match.com over different discrete time intervals.
- It is important to note there are many users that just browse and do not interact with the other users on the site. They were excluded them from the data set.
  - If a user had an exchange with another user, then and only then they were counted as UU, otherwise considered to be passive users.
  - These users have little to no effect on the site, hence are not incorporated into the model.

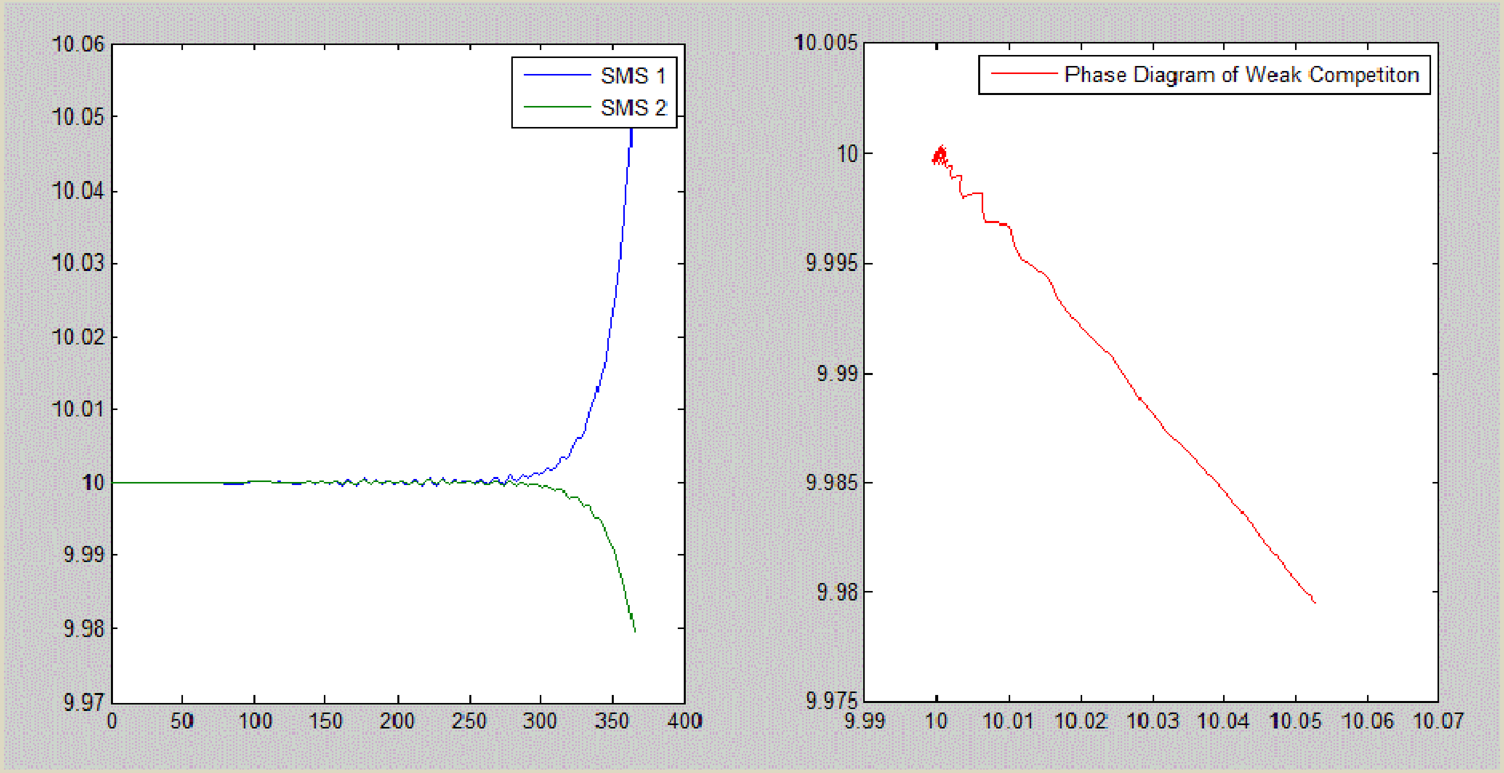
Analysis

- An OLS analysis was conducted on the plot of the data to determine the best fit.
  - The set of data was parameter fitted to a regression model using Science Workshop, a Microsoft Excel based program that runs various OLS regression models.
- The two sets of data were plotted individually and run through the parameter fitting process.
  - From this the intrinsic rate, internal interaction rate and the external interaction rate were determined.

Analysis (The Parameters)

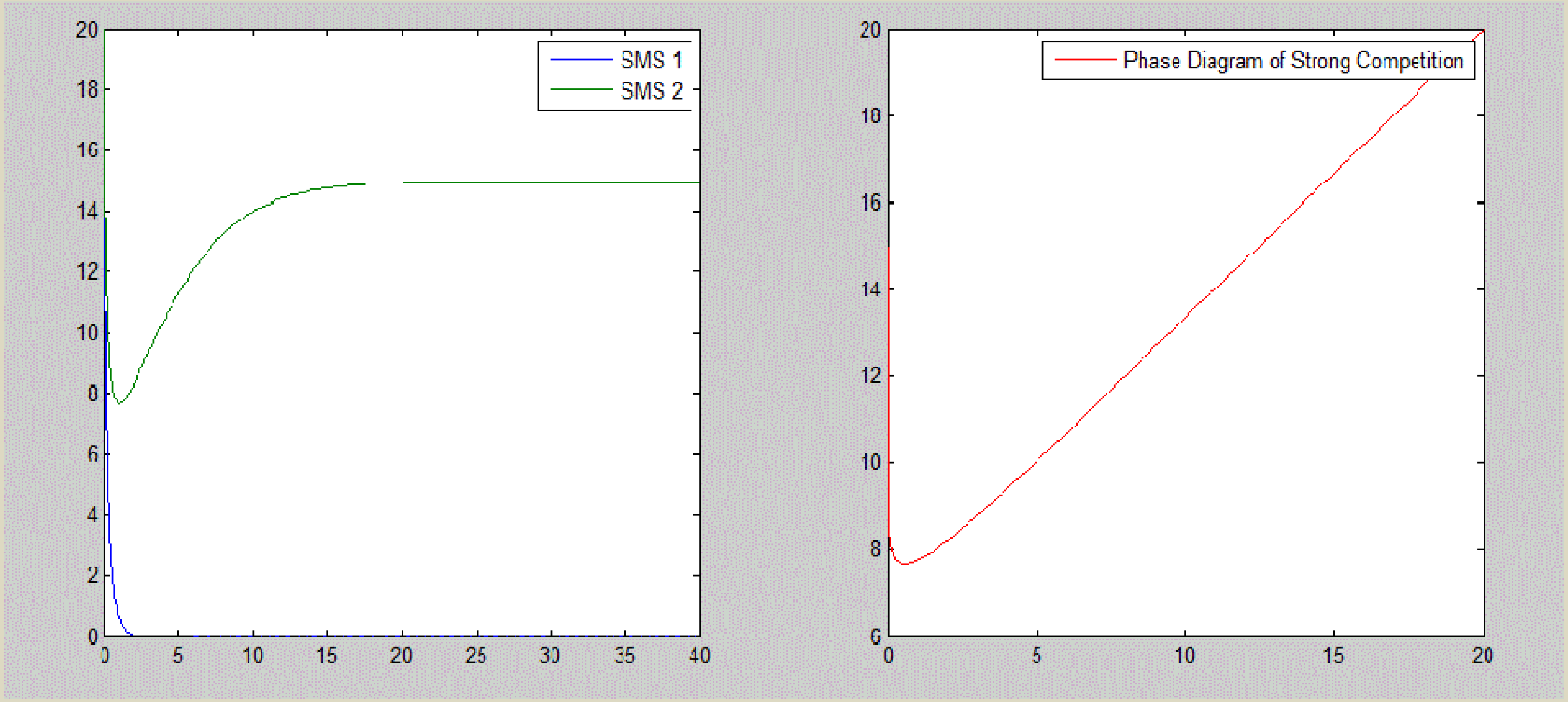
- Plot the two data sets individually and then running the OLS by using the ScienceWorkshop software program.
  - The equilibrium states were determined for each plots
- The Multi-Variable Fit program then utilizes the initial, overall rate of growth , and equilibrium values to estimate the parameters for each set of data.

The Modeling Results: Weak External Interaction



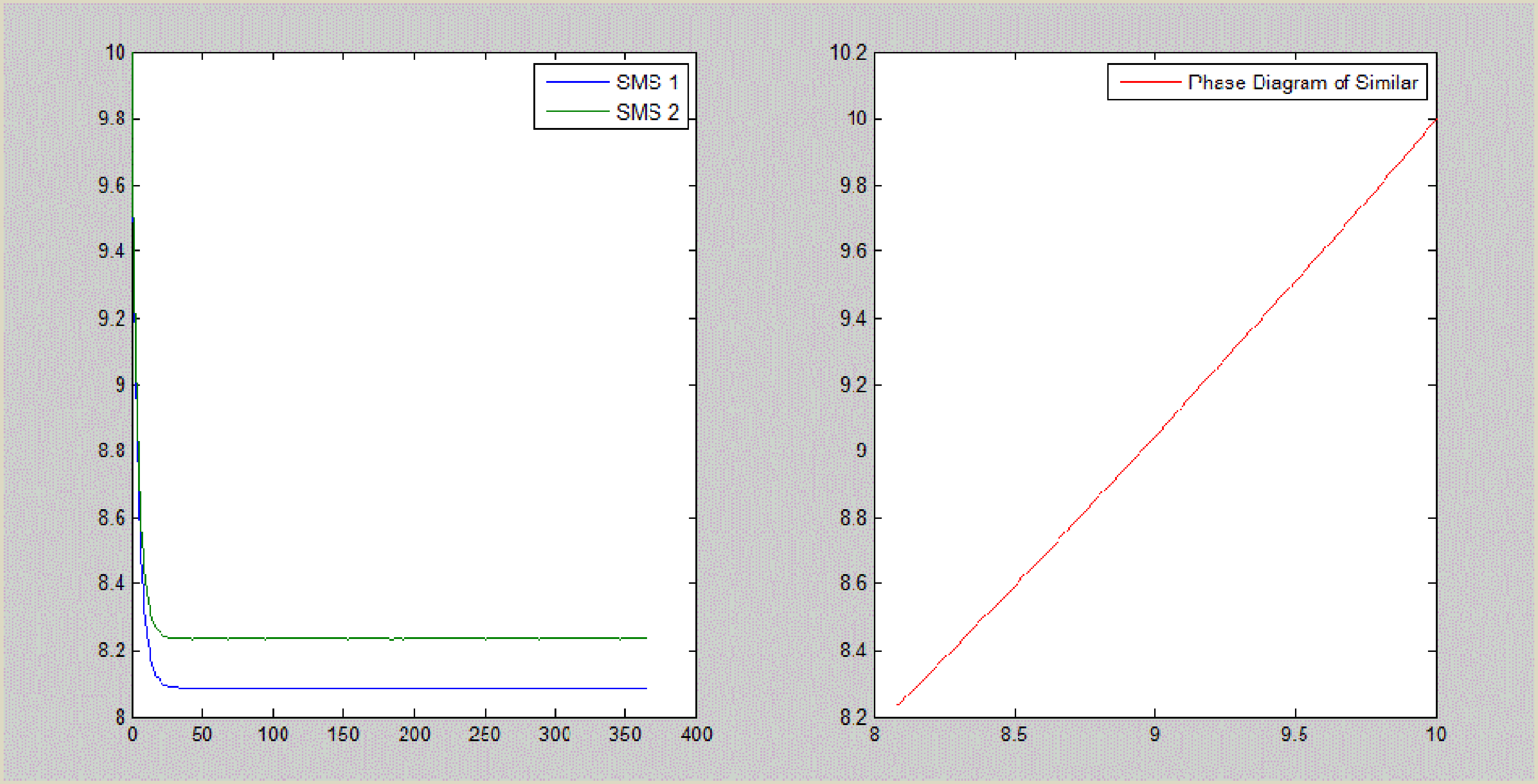
$a_x = 0.5$, $b_x = -0.01$, $c_{xy} = -0.04$, $a_y = 0.3$, $b_y = -0.02$, $c_{yx} = -0.01$

The Modeling Results: Strong External Interaction



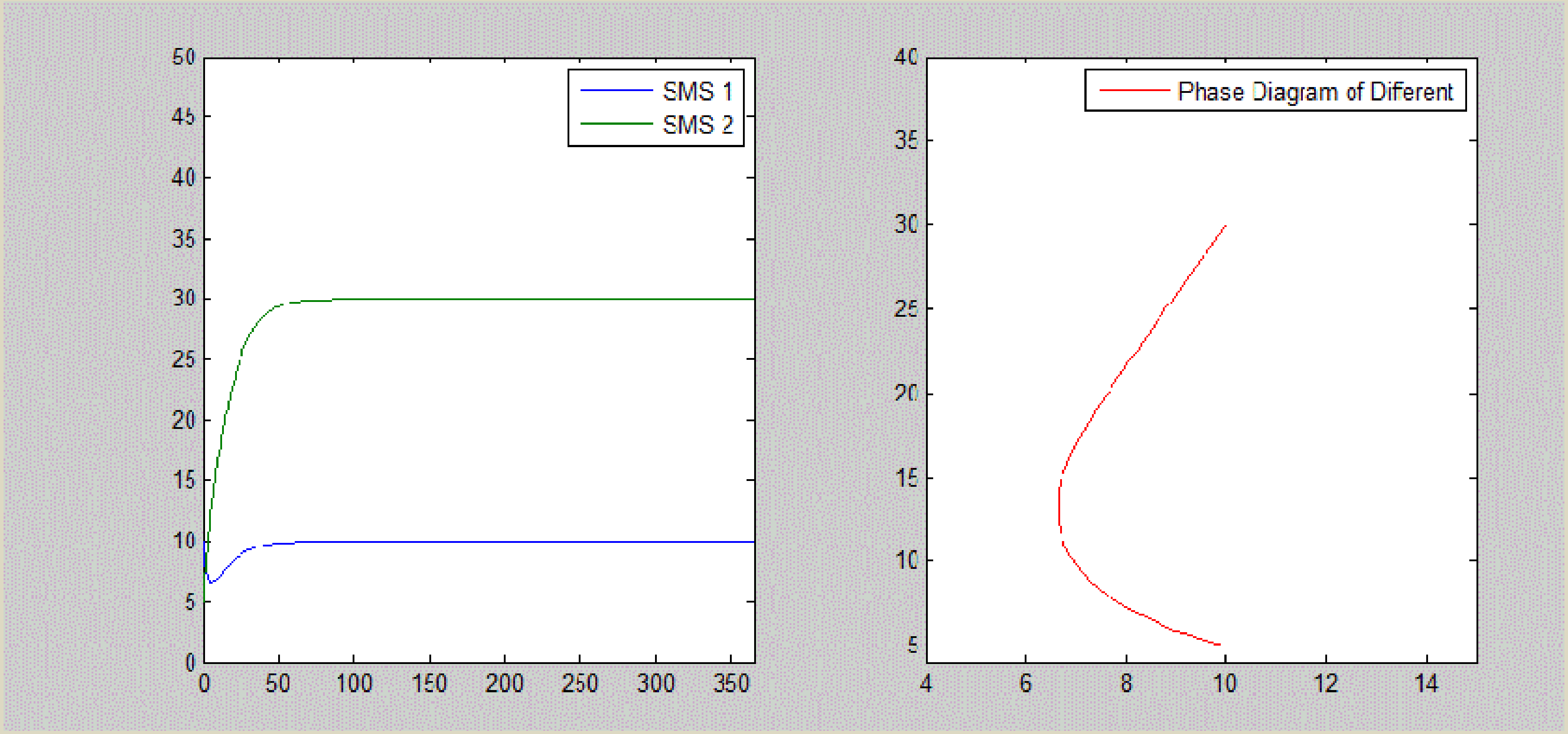
$a_x = 0.5$, $b_x = -0.01$, $c_{xy} = -0.4$, $a_y = 0.3$, $b_y = -0.02$, $c_{yx} = -0.25$

The Modeling Results: Similar Niche Role



$a_x = 0.2$, $b_x = -0.04$, $c_{xy} = 0.015$, $a_y = 0.25$, $b_y = -0.05$, $c_{yx} = 0.2$

The Modeling Results:
Different Niche Role

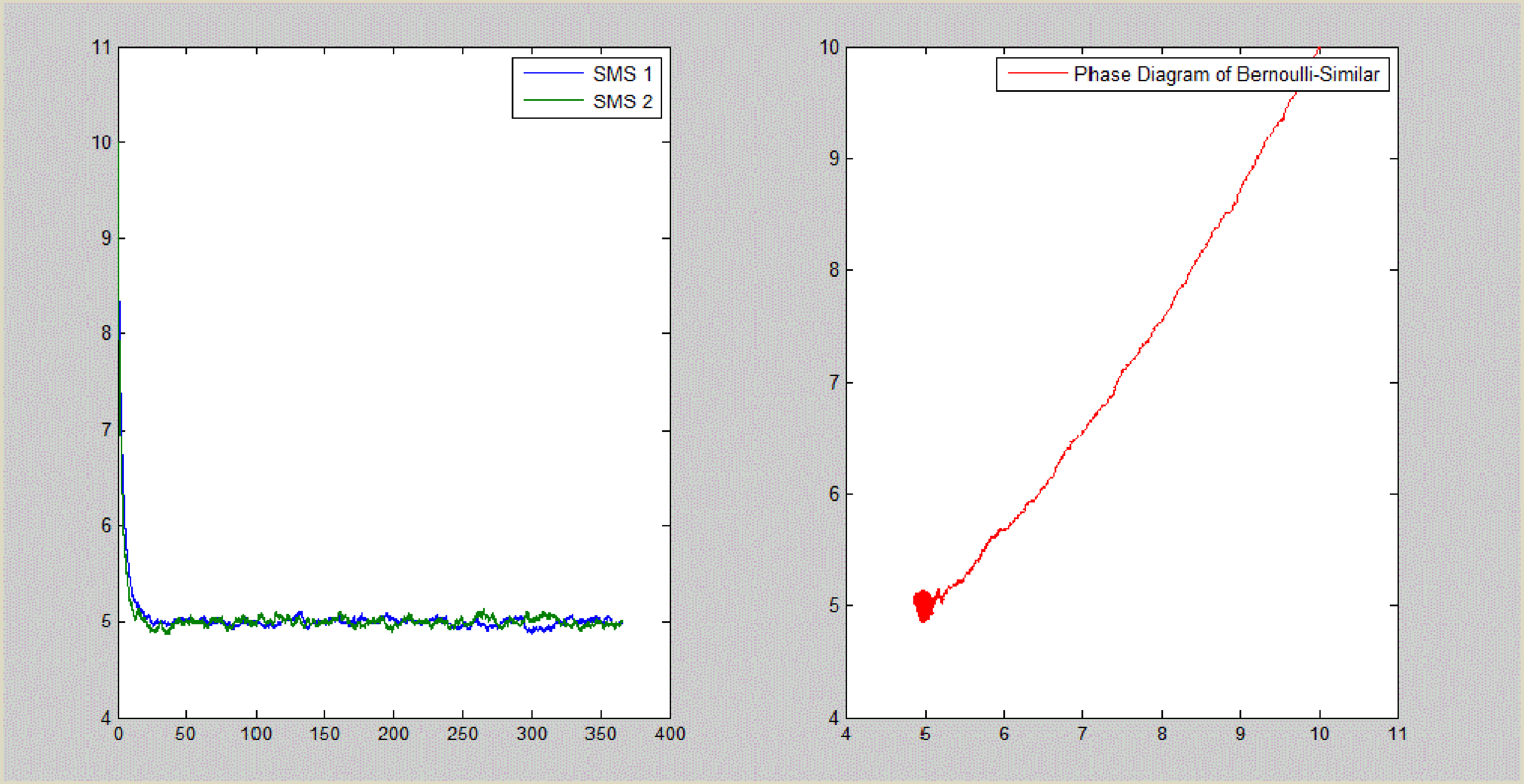


$a_x = 0.2$, b_x=-0.05, c_{xy}=0.01, a_y=-0.05, b_y=-0.015, c_{yx}= 0.05

Initial Results

- Gives results with a given set of data, hence fixed coefficients.
- However, in real world situations the parameters are not set and actually evolve over time.
  - The interactions are stochastic processes.
- Further model the interactions when the external interaction rates are biased random-walks. C_{ij} changes values at each time step.
  - Will examine Different and Similar Competition Models
    - a_i and b_i are the same as those in the previous model.

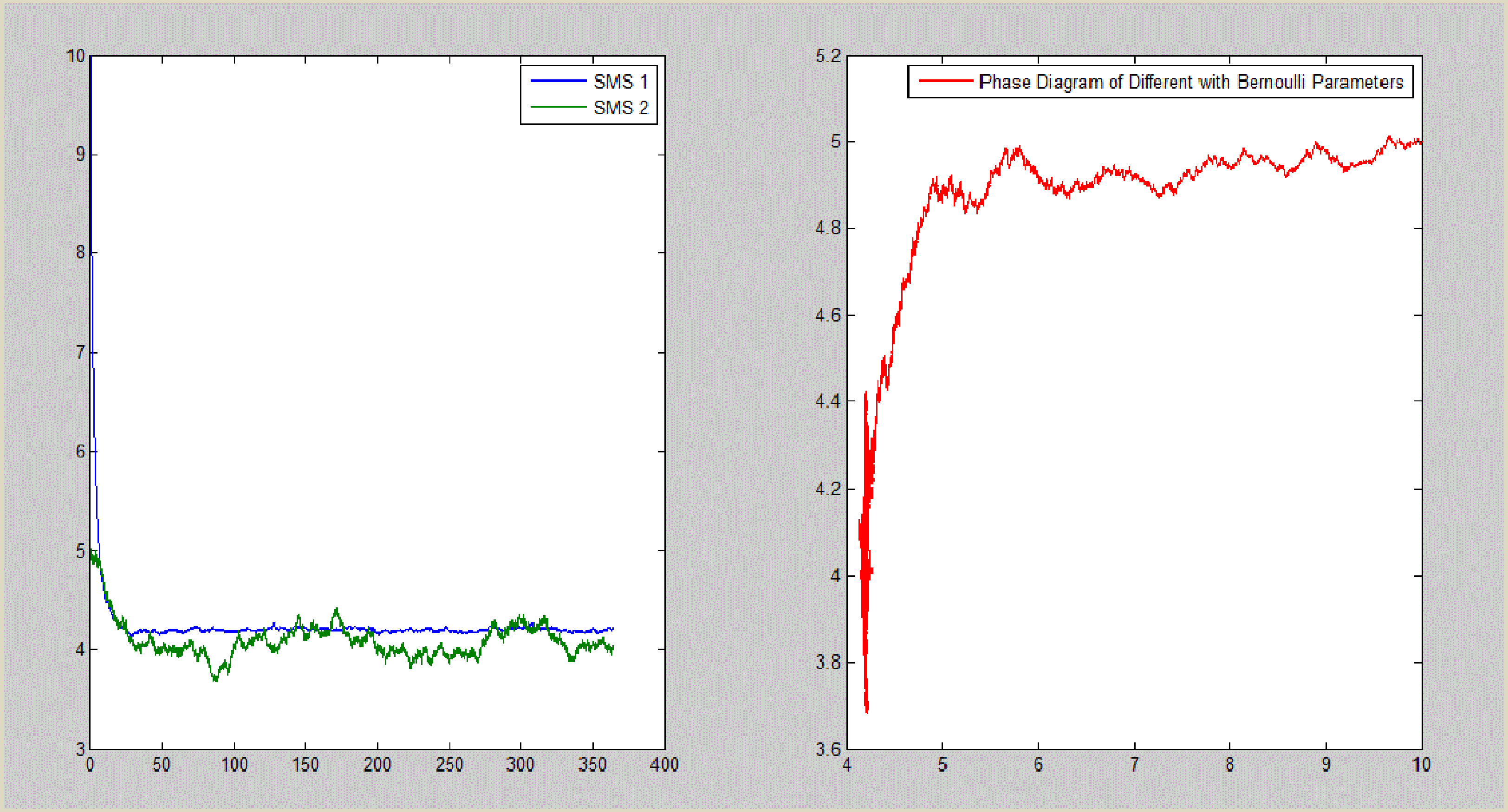
The Modeling Results: Similar Niche Role, with Stochastic c_{ij}



$$c_{xy} = \begin{cases} 0.015, & \\ -0.015, & p = .5 \end{cases}$$

$$c_{yx} = \begin{cases} 0.02, & \\ -0.02, & p = .5 \end{cases}$$

The Modeling Results:
Different Niche Role, with Stochastic c_{ij}



$$c_{xy} = \begin{cases} 0.015, & \\ -0.01, & \end{cases} \quad p = .5$$

$$c_{yx} = \begin{cases} 0.055, & \\ -0.05, & \end{cases} \quad p = .5$$

The Substantive Results

- The interactions between two social media sites can be modeled by Lotka-Volterra equations, by representing the external interaction rate as a stochastic process.
- The external interaction rate serves as a key driver for the growth of Unique Users between two Social Media Sites.
- The model more closely mimics a real word situation where the interactions between two sites is constantly evolving over time.
- Examples of this include match.com, who to compete with eHarmony added their own match-making algorithm to attract users that were looking for life-partners.
  - Even though eHarmony and match.com are in the same general social media category, each plays a different niche role. (Match.com is for general dating, while eHarmony.com is for more serious singles looking for life partners).
- A more recent example is Facebook constantly eliminating their competitors through acquisitions (Instagram) or Facebook continually offering new technologies to users (Timeline).

Moving Forward

- Future work:
  - Model growth with general intrinsic and internal interaction rates. Not fixing the coefficients but allowing them to adapt and evolve in the system.
  - Allow for correlation between the interaction rates. This will more closely model real world situations in which user networks are intertwined.
    - Incorporate preference into the system (Indian Buffet Process).
  - Increase number of populations and interpret how the driving processes effect the growth of the Social Media Sites.
  - Allow for users to join multiple sites (multi-homing), as the literature shows that 40% of people maintain multiple profiles over various sites. (Lenhart 2009)

Bibliography

A. Castiaux, “Inter-organisational learning –Lotka-Volterra modelling of different types of relationships” International System Dynamics Conference – Oxford, July 2004

Banerji, A., B. Dutta. 2009. Local Network Externalities and Market Segmentation. *International Journal of Industrial Organization* **27(5)** 605–614.

Epstein, Joshua M. *Nonlinear Dynamics, Mathematical Biology, and Social Science. Reading, MA: Addison-Wesley Pub., 1997. Print.*

Huang, Y., P. Singh, A. Ghose. 2011. A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media. Working Paper.

Morris, Steven A., and David Pratt. "Analysis of the Lotka–Volterra Competition Equations as a Technological Substitution Model." *Technological Forecasting & Social Change 70.2 (2011): 103-33. Print.*

Trusov, M., R.E. Bucklin, K. Pauwels. 2009. Effects of Word-of-mouth Versus Traditional Mar- keting: findings from an Internet social networking site. *Journal of Marketing* **73(5)** 90–102.

Zhang, Kaifu, and Miklos Sarvary. *Social Media Competition: Differentiation with User-Generated Content. 1 Sept. 2011. Web.*

http://siteanalytics.compete.com/match.com/

http://siteanalytics.compete.com/eharmony.com/