Social epidemiology and intimate social networks

Courtney D. Corley*, Susie Ramisetty-Mikler†

*Computational Epidemiology Research Laboratory (CERL), Department of Computer Science and Engineering, University of North Texas, Denton, TX USA
†School of Public Health, University of Texas Health Science Center, Dallas, TX USA

November 6, 2007

http://cerl.unt.edu/~corley
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
Social Epidemiology addresses the links between behavioral, psycho-social risk factors, as well as the protective effects of social networks and support as they relate to disease and health of populations.

Social networks link macro social forces
i.e. income distribution, social cohesion and social capital and inequalities in socioeconomic status, and residential segregation – to patterns of health and disease in populations.

Social networks form in dynamic environments

- constant change in the population composition in terms of: age, gender, and ethnicity
- population features, such as socio-economic levels, are proven to correlate with disease impact severity.
- Individual, community and population-level behavior variations
Networks you may have heard about

Natural and man-made
- Internet topology
- U.S. Interstate system
- Cortical brain pathways

Social / human interaction
- Scientific collaboration, Enron e-mails
- Web 2.0 and Internet communities:
  - MySpace, Facebook, LinkedIn, LiveJournal

Romantic and intimate interactions
- one hopes to engage in this network type
Survey measures as sources of social networks

Common *patterns* of networks can be *observed* from the findings of general population surveys:

- Correlating surveyed risk behaviors and disease surveillance
- Common survey data:
  - age, gender, ethnicity, education, income / socio-economic status, employment (status and type of occupation), health insurance status, geographic location (census block or "rural/urban")
- Geography is a core element of the U.S. Census Bureau’s system for organizing and presenting statistical data.
- Census Bureau geographic entities:
  - state, county, census tract, block group, and census block
Geographic Hierarchy for U.S. Census

Inset A: MA Components

In 44 States, DC, and Puerto Rico

- MAs
- Counties
- Census Tracts/BNA
- Block Groups
- Blocks

In the Six New England States

- MAs
- Cities
- Minor Civil Divisions
  - (Towns)
- Census Tracts/BNA
  - Block Groups
  - Blocks

CD Corley and S Ramisetty-Mikler
EPIGRID November 6, 2007
United States general population surveys

General Population
U.S. National Surveys

Alcohol and Drug Surveys
- National Survey on Drug Use and Health (NSDUH)

Risk Behavior Surveys
- Youth Risk Behaviors Survey (YRBS)
- Adult Behavioral Risk Factor Survey (BRFSS)

Nutrition and Health Surveys
- National Health Interview Survey (NHIS)
- National Health and Nutrition Examination Survey (NHANES)
Examples of gen. pop. surveys in the United States

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Number of Objectives Tracked</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health Interview Survey (NHIS)</td>
<td>67</td>
</tr>
<tr>
<td>National Health and Nutrition Examination Survey (NHANES)</td>
<td>35</td>
</tr>
<tr>
<td>National Vital Statistics System—Mortality (NVSS-M)</td>
<td>32</td>
</tr>
<tr>
<td>National Survey of Family Growth (NSFG)</td>
<td>14</td>
</tr>
<tr>
<td>National Hospital Discharge Survey (NHDS)</td>
<td>11</td>
</tr>
<tr>
<td>Youth Risk Behavior Surveillance System (YRBSS)</td>
<td>11</td>
</tr>
<tr>
<td>HIV/AIDS Surveillance System</td>
<td>10</td>
</tr>
<tr>
<td>Behavioral Risk Factor Surveillance System (BRFSS)</td>
<td>9</td>
</tr>
<tr>
<td>National Household Survey on Drug Abuse (NHSDA)</td>
<td>8</td>
</tr>
<tr>
<td>School Health Policies and Programs Study (SHPSS)</td>
<td>8</td>
</tr>
<tr>
<td>National Vital Statistics System—Nativity (NVSS-N)</td>
<td>3</td>
</tr>
<tr>
<td>National Profile of Local Health Departments (NFLHD)</td>
<td>3</td>
</tr>
<tr>
<td>National Ambulatory Medical Care Survey (NAMCS)</td>
<td>7</td>
</tr>
<tr>
<td>United States Renal Data System (USRDS)</td>
<td>7</td>
</tr>
<tr>
<td>STD Surveillance System (STDS)</td>
<td>7</td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey (MEPS)</td>
<td>6</td>
</tr>
<tr>
<td>National Hospital Ambulatory Medical Care Survey (NHAMCS)</td>
<td>6</td>
</tr>
<tr>
<td>Continuing Survey of Food Intake by Individuals (CSFII)</td>
<td>6</td>
</tr>
<tr>
<td>National Crime Victimization Survey (NCVS)</td>
<td>6</td>
</tr>
<tr>
<td>1999 National Worksite Health Promotion Survey (NWHP5)</td>
<td>5</td>
</tr>
<tr>
<td>State Tobacco Activities Tracking and Evaluation System (STATES)</td>
<td>5</td>
</tr>
<tr>
<td>National Notifiable Disease Surveillance System (NNDS)</td>
<td>5</td>
</tr>
<tr>
<td>Monitoring the Future Study (MTF)</td>
<td>5</td>
</tr>
</tbody>
</table>
Implications for population surveys: an example

Youth Risk Behavior Surveillance System (YRBSS)

- Nationally representative sample of students in grades 9 – 12 in the U.S.

- The national YRBS uses a three-stage, cluster sample design to obtain a nationally representative sample.
  1) primary sampling units (PSUs) consisting of large-sized counties or groups of smaller, adjacent counties.
  2) Selection of schools: whole schools (9-12) or cluster schools; large (> 25) or small (< 25); $\frac{1}{4}$ of PSUs are small.
  3) Selection of one or two entire classes in each chosen school and in each of grades 9–12.

- The target population comprises all public and private high school students in the 50 states and the District of Columbia.
YRBSS data features
But why?

Create knowledge-database of impact-methodologies

- *impacts*: global, community, individual, economic and pathogen
  - Vaccination (HPV, HBV)
  - Information diffusion
  - Behavior change

- Population demographic dynamics
- Social-affinity intimate network formation processes
- Community structure, clustering, centrality
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
Selected social network and epidemic models

- Multi-agent based simulation
  - EPISIMS: U. of Maryland, Virginia Tech, Los Alamos
    (Eubank, Guclu, Kumar, Marathe, Srinivasan, Toroczkai & Wang 2004)
  - BIOWAR: Carnegie Melon University (?)

- Mean-field and ODE-type
  - (Hughes, Garnett & Koutsky 2002)
  - (Sanders & Taira 2003)
  - (Corley & Mikler 2005)

\[
\begin{align*}
\Delta S_{kl} &= 0.5\mu(1 - \phi_{kl})\sum_{i=1}^{\omega_i} \omega_i \eta - (\lambda_{kl} + \nu)S_{kl} + \sigma V_{kl} \\
\Delta I_{kl} &= \lambda_{kl}S_{kl} - (\gamma + \nu)I_{kl} \\
\Delta R_{kl} &= \gamma I_{kl} + \alpha \gamma V I_{kl} - \nu R_{kl} \\
\Delta V_{kl} &= 0.5\mu\phi_{kl} \sum_{i=1}^{\omega_i} \omega_i \eta - (\nu + \sigma + \psi \lambda_{kl})V_{kl} \\
\Delta V I_{kl} &= \psi \lambda_{kl} V_{kl} - (\nu + \alpha \gamma) V I_{kl}
\end{align*}
\]
A Swedish survey on sexual behavior was analyzed and reported by Liljeros et al. in a 2001 Nature article

- random sample of 4,781 Swedes
- ages 18-74
- Example question: **how many intimate partner changes occurred in a year's time?**
  - probability distribution for having \( k \) intimate partners.
  - number of partners in the previous year follows a *power-law* distribution.

- Cumulative probability function \( p(k) \approx L(k)k^{-\alpha} \)
  - scaling parameter \( \alpha > 1 \) and \( L(k) \) being a slowly varying function that controls the shape and finite extent of the lower tail

- **Preferential attachment** can describe *power-law* degree distribution.
  - results when a new node is more likely to connect to a node with a high degree than to a node of low degree.
  - \[ \Pi_j = \frac{d_j^o}{\sum_{v_i \in G_k} d_{v_i}^o} \]

1 Introduction
   - Social Epidemiology
   - Previous Work

2 Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3 Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
Classical graph statistics

Consider $G = (V, E)$

Neighborhood:
$$N(v) = \{ \{u\} : e_{u,v} \in E\}$$

Size:
$$n = |V|, m = |E|.$$

Degree:
$$d^o_v = |N(v)|.$$

Average degree:
$$k = \frac{2m}{n}.$$

Density: Represents the probability any two randomly chosen vertices are connected,
$$\delta(G) = \frac{2m}{n(n-1)}.$$

Degree distribution:
$$p(k) = \frac{d^o_v}{\sum_{v \in V} d^o_v = k}.$$

Cumulative degree distribution:
$$P_k = \sum_{k' = k}^\infty p_{k'}.$$

Clustering coefficient: The clustering coefficient $C_v$ is the probability that any two nodes are linked together if they have a neighbor in common.
$$C_v = \frac{|E_{N(v)}|}{|N(v)|(|N(v)|-1) / 2} = \frac{2|\{e(y,u)\}|}{d^o_v(d^o_v-1)} : y, u \in N(v), e(y,u) \in E.$$

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i.$$

(Erdos & Renyi 1959, Newman & Park 2003, Watts & Strogatz 393)
Consider \( G = (\top, \bot, E) \) and \( G^* = (\top \cup \bot, E) \).

- **Size**: \( n_\top = |\top|, \ n_\bot = |\bot| \),

- **Average degree**
  - \( k_\top = \frac{m}{n_\top}, \ k_\bot = \frac{m}{n_\bot} \).
  - \( k = \frac{2m}{n_\top + n_\bot} = \frac{n_\top k_\top + n_\bot k_\bot}{n_\top + n_\bot} \).

- **Density**: \( \delta(G^*) \ll \delta(G) \)
  - \( \delta(G) = \frac{m}{n_\top n_\bot} \).
  - \( \delta(G^*) = \frac{2m}{(n_\top + n_\bot)(n_\top + n_\bot - 1)} \).

- **Clustering coefficient** consider a pair of nodes, both in either \( \top \) or \( \bot \)
  - \( cc_\bullet \) captures the overlap in neighborhoods of vertices \( u \) and \( v \).
  - \( cc_\bullet(u, v) = 0 \)
    - The neighborhoods of vertices \( u \) and \( v \) do not overlap.
  - \( cc_\bullet(u, v) = 1 \)
    - \( u \) and \( v \) are elements of the same neighborhood.
Bipartite clustering: an example

The neighborhoods of vertices $u$ and $v$ intersect at vertices $a$ and $b$.

$$\text{cc}_\bullet(u, v) = \frac{2}{3}$$

However, there is no overlap (clustering) between vertices $u$ and $w$, clustering coefficient of vertex $u$ remains the same

$$\text{cc}_\bullet(u) = \frac{2}{3}$$

- $\text{cc}_\bullet(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$
- $\text{cc}_\bullet(u) = \frac{\sum_{v \in N(N(u))} \text{cc}_\bullet(u, v)}{|N(N(u))|}$
- $\text{cc}_\bullet(G) = \frac{n_\top \text{cc}_\bullet(\top) + n_\perp \text{cc}_\bullet(\perp)}{n_\top + n_\perp}$
- $\text{cc}_\uparrow(u, v) = \frac{|N(u) \cap N(v)|}{\min(|N(u)|, |N(v)|)}$
- $\text{cc}_\downarrow(u, v) = \frac{|N(u) \cap N(v)|}{\max(|N(u)|, |N(v)|)}$
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
We have developed a simulator capable of building an evolving social network of dynamic heterosexual intimate contacts mapped to:

bipartite graph described by the triplet
\[ G = (G_f, G_m, \vec{E}) \]
- \( G_M \) represents male vertices in network at any temporal step
- \( G_F \) represents female vertices in network at any temporal step
- \( \{ \vec{E} : \vec{E} \subseteq G_f \times G_m \} \) is a vector containing the set of edges present during discrete temporal intervals.

socio-demographic feature vector for a vertex with gender \( k \) (\( v_k \)):
\[ \vec{D}_{v_k} \]

the affinity (pairwise socio-feature similarity score) between two vertices \( v_m \) and \( v_f \):
\[ p_{v_m,v_f}(mixing) \]

the probability that edge \( E \) will be placed in the graph:
\[ p_E(Attach) \]
DynSNIC general algorithm

begin
input : user defined parameter space
Insert $n_m$ vertices in $G_m$
Insert $n_f$ vertices in $G_f$
foreach $v_f \in G_f$ do
  \[ \text{maxDegree}_{vf} \leftarrow \text{zipf}(-2.54, kFemaleBound) \]
  draw a random vector $\tilde{r}$ from a user defined and bounded probability distribution
  \[ \vec{D}_{vf} \leftarrow \tilde{r} //\text{demo. property vector stoch. created and assigned to female vertex} \]
foreach $v_m \in G_m$ do
  \[ \text{maxDegree}_{vm} \leftarrow \text{zipf}(-2.31, kMaleBound) \]
  draw a random vector $\tilde{r}$ from a user defined and bounded probability distribution
  \[ \vec{D}_{vm} \leftarrow \tilde{r} //\text{demo. property vector stoch. created and assigned to male vertex} \]
\forall v_f \in G_f \text{ calculate preferential attachment } p(k)
\forall v_m \in G_m \text{ calculate preferential attachment } p(k)
foreach \text{discrete time step} do
  \[ E_t = \{\} //\text{initialize current edge set to the empty set} \]
  \[ \text{optBiMatching}(G') //\text{Optimize bipartite matching on dynamic network, see Alg.2} \]
  Evaluate disease dynamics including any intervention strategies
foreach $v_n \in G'$ do
  //population evolution
  draw a uniformly distributed random number $r$
  if \[ r > p(\text{aging} - \text{out}) \] then
    color $v_n \in G$ as unusable
    create new node $v_{\text{new}}$
    \[ \text{maxDegree}_{v_{\text{new}}} \leftarrow \text{zipf}(\alpha, \text{genderBound}) \]
    insert $v_{\text{new}}$ in $G_k$
    recalculate preferential attachment $p(k) \forall v_k \in G'$
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
Cartoon population – $k$ partners

Minimal remaining degree
Every vertex has one connection:
only consider sexually active population
Cartoon example: "optimal configuration"
Exhausting total degree, set by set
Exhausting total degree, vertex by vertex
Optimal matching heuristic: vertex by set
Optimal matching heuristic: vertex by set

Diagram: A graph with vertices and edges illustrating the matching process.
Optimal matching heuristic: vertex by set
Optimal matching heuristic: vertex by set
Optimal matching heuristic: vertex by set
**DynSNIC** takes a coarse first cut at scoring the likelihood of mixing between two individuals, using the unweighted cosine similarity of both vertices’s demographic feature vector.

\[
p_{v_m,v_f}(mixing) = \text{COSIM}(\vec{D}_v, \vec{D}_m) = \cos \theta = \frac{\vec{D}_v \cdot \vec{D}_m}{|\vec{D}_v||\vec{D}_m|}
\]

\[
p_E(Attach) = p_{v_m}(k) \times p_{v_m,v_f}(mixing)
\]
Connecting a bipartite graph minimizing remaining degree

\[ \text{optBiMatching}(G') \text{ where } G' = \{ G_{m\text{usable}}, G_{f\text{usable}}, E_t \} \]

\begin{verbatim}
input : G = (G_f, G_m, E) where E = ∅ 
output: Maximally connected bipartite graph of intimate contacts 

Begin 
while \( \exists v_m \in G_m \) and \( \exists v_f \in G_f \) s.t. \( d^o_{v_k} < \text{maxDegree}_{v_k} \) do 
  in vertex order given \( v_{kid} \) 
  choose \( v_m \in G_m \) s.t. \( d^o_{v_m} < \text{maxDegree}_{v_m} \) 
  loopCount = 0 
  maxReached = FALSE 
  repeat 
    inserted = FALSE 
    loopCount++ 
    randomly choose \( v_f \in G_f \) s.t. \( d^o_{v_f} < \text{maxDegree}_{v_f} \) 
    \( p_E(\text{Attach}) = p_{v_f}(k) \times p_{v_m, v_f}(\text{mixing}) \) 
    draw a uniformly-distributed random number \( r \) 
    if \( p_E(\text{Attach}) > r \) then 
      add \( E = (v_f, v_m) \) in \( G \) 
      inserted = TRUE 
    if !inserted and maxLoopsReached then 
      arbitrarily choose \( v_f \in G_f \) s.t. \( d^o_{v_f} < \text{maxDegree}_{v_f} \) 
      add \( E = (v_f, v_m) \) in \( G \) 
      inserted = TRUE 

  in vertex order given \( v_{kid} \) 
  until inserted == TRUE 
// continued on next slide 
\end{verbatim}
choose $v_f \in G_f$ s.t. $d_{v_f}^o < \text{maxDegree}_{v_f}$

loopCount = 0
maxReached = FALSE

repeat

inserted = FALSE
loopCount++

randomly choose $v_m \in G_m$ s.t. $d_{v_m}^o < \text{maxDegree}_{v_m}$

$p_E(\text{Attach}) = p_{v_m}(k) \times p_{v_m,v_f}(\text{mixing})$

draw a uniformly-distributed random number $r$

if $p_E(\text{Attach}) > r$ then

add $E = (v_f, v_m)$ in $G$

inserted = TRUE

if !inserted and maxLoopsReached then

arbitrarily choose $v_m \in G_m$ s.t. $d_{v_m}^o < \text{maxDegree}_{v_m}$

add $E = (v_f, v_m)$ in $G$

inserted = TRUE

until inserted == TRUE
A DynSNIC realization: 100 node ($|G_m| = |G_f|$)
Outline

1. Introduction
   - Social Epidemiology
   - Previous Work

2. Dynamic Evolving Intimate Contact Social Networks: DynSNIC
   - Concept review: graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3. Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination: A sample case study
10,000 vertices ($|G_m| = |G_f|$) 
10 runs - 10 realizations each run 
Cumulative intimate contact degree distribution
DynSNIC’s clustering statistics by affinity scoring method

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$p(k)$</th>
<th>COSIM</th>
<th>$p(k) \times$ COSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>7693</td>
<td>7716</td>
<td>7742</td>
</tr>
<tr>
<td>$\delta(G)$</td>
<td>3.08E-4</td>
<td>3.09E-4</td>
<td>3.10E-4</td>
</tr>
<tr>
<td>$\delta(G^*)$</td>
<td>1.54E-4</td>
<td>1.54E-4</td>
<td>1.55E-4</td>
</tr>
<tr>
<td>$cc(Males)$</td>
<td>0.423</td>
<td>0.457</td>
<td>0.442</td>
</tr>
<tr>
<td>$cc(Females)$</td>
<td>0.218</td>
<td>0.250</td>
<td>0.233</td>
</tr>
<tr>
<td>$cc(G)$</td>
<td>0.321</td>
<td>0.353</td>
<td>0.337</td>
</tr>
<tr>
<td>$cc_{\uparrow}(Males)$</td>
<td>0.786</td>
<td>0.810</td>
<td>0.799</td>
</tr>
<tr>
<td>$cc_{\uparrow}(Females)$</td>
<td>0.848</td>
<td>0.861</td>
<td>0.853</td>
</tr>
<tr>
<td>$cc_{\uparrow}(G)$</td>
<td>0.817</td>
<td>0.835</td>
<td>0.826</td>
</tr>
<tr>
<td>$cc_{\downarrow}(Males)$</td>
<td>0.456</td>
<td>0.486</td>
<td>0.470</td>
</tr>
<tr>
<td>$cc_{\downarrow}(Females)$</td>
<td>0.226</td>
<td>0.255</td>
<td>0.241</td>
</tr>
<tr>
<td>$cc_{\downarrow}(G)$</td>
<td>0.321</td>
<td>0.370</td>
<td>0.337</td>
</tr>
</tbody>
</table>
Outline

1 Introduction
   - Social Epidemiology
   - Previous Work

2 Dynamic Evolving Intimate Contact Social Networks : DynSNIC
   - Concept review : graph metrics
   - DynSNIC general algorithm
   - Optimal bipartite matching

3 Experimental Results
   - Monte-Carlo simulations
   - HPV Vaccination : A sample case study
Many Human Papilloma Virus (HPV) types are sexually transmitted and HPV DNA is found in 99.7% of all cervical cancers with HPV-types 16, 18, 31 and 45 accounting for 75% of cervical dysplasia (Goldie, Kohli & Grima 2004).

- **natural infection**
  - chance of infection in one encounter $p_{ik}$
  - number of encounters ($\lambda$)
  - $p_n(i) = 1 - [1 - p_{ik}]^\lambda$

- **breakthrough infection**
  - combines intervention efficacy $e_{int}$
  - chance of natural infection
  - $p_b(i) = e_{int} \times p_{ik}$

- **biological parameters**
  - probability of acquiring HPV in one encounter (0.08 male-to-female, 0.02 female-to-male)
  - encounter frequency drawn from a Poisson distribution with a mean of 50
  - intervention efficacy is 75%
  - the age-range modeled is 50 years
  - infection clears after two years

5% of the population is initially infected.
Prophylactic vaccination impact

HPV infectious population per gender and three intervention solutions

10,000 pop. size, 10 Monte-Carlo simulations, 30 realizations each run $|G_m| = |G_f|$
RRP and related models

- DynSNIC: 0.75 RRP*
- Hughes et al cite a RRP of 0.68 with a range of 0.628 to 0.734†
- Sanders and Taira cite a RRP of 0.8 and above‡

*relative reduction in prevalence.
†(Hughes et al. 2002) ‡(Sanders & Taira 2003)
Introduction
Dynamic Evolving Intimate Contact Social Networks: DynSNIC

Experimental Results

Monte-Carlo simulations
HPV Vaccination: A sample case study


