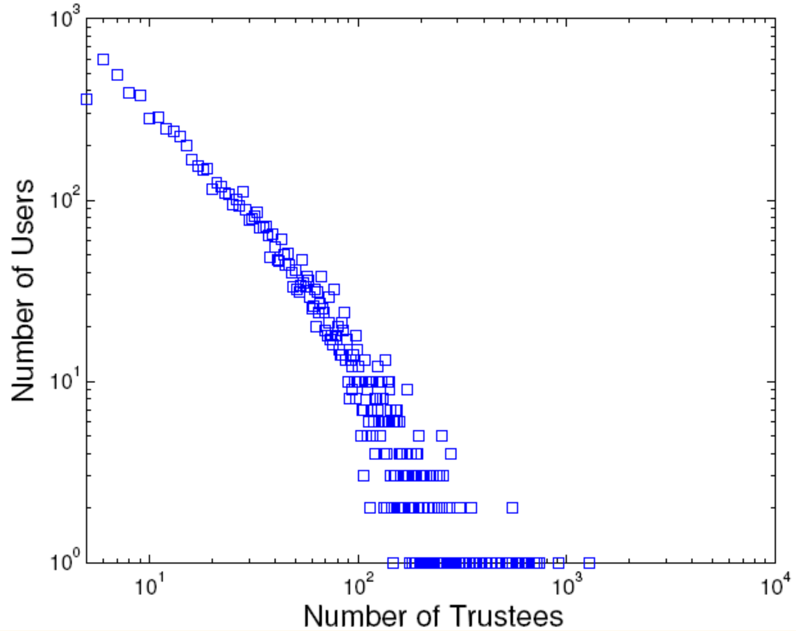
Exploiting Homophily Effect for Trust Prediction

Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu
Computer Science and Engineering
Arizona State University

February 4-8, 2013 WSDM2013

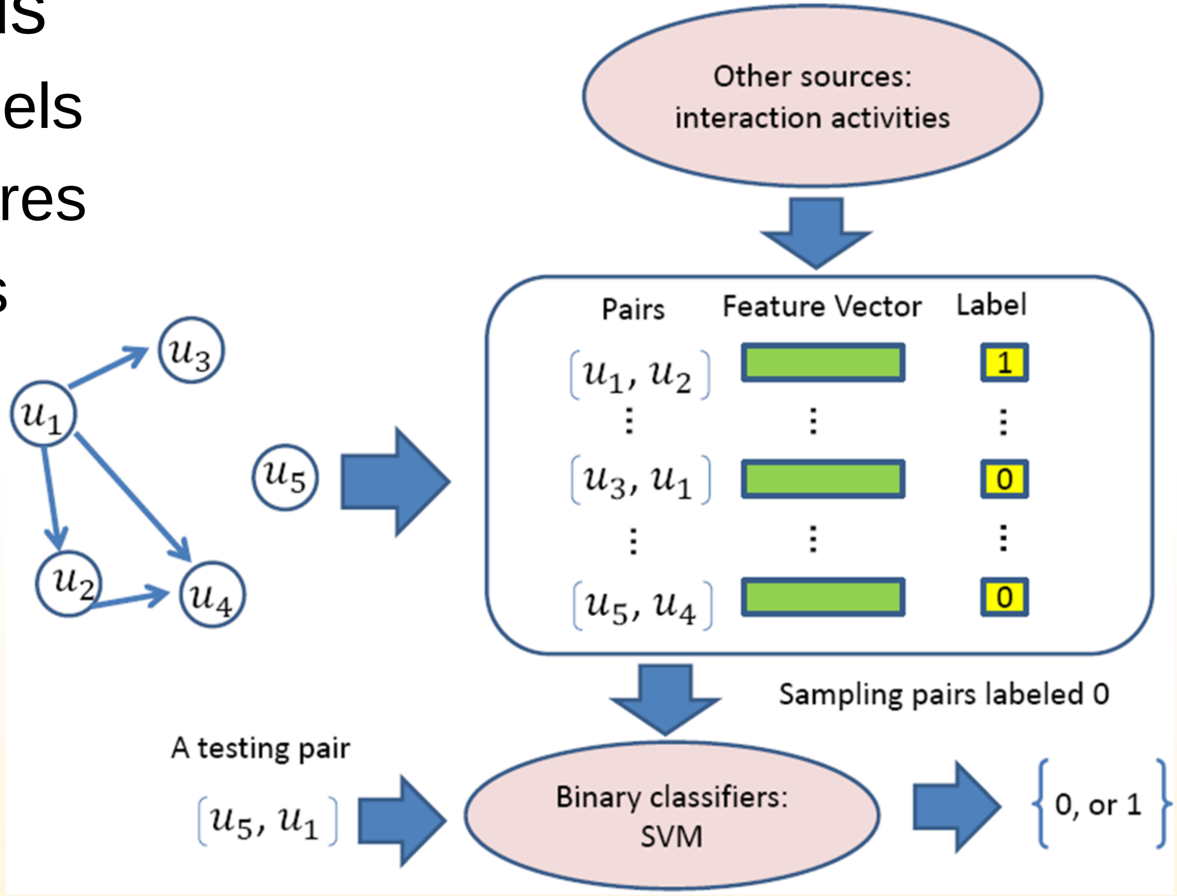
Online Trust

- Trust plays an important role in helping online users collect reliable information for decision making
- The available explicit trust relations are extremely sparse



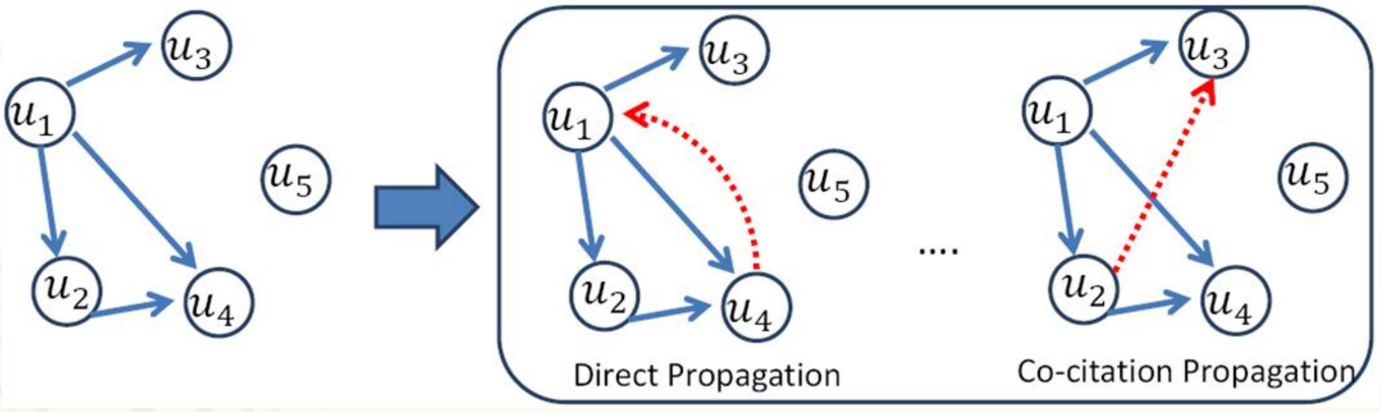
Existing Trust Predictors (Supervised)

- Supervised methods
  - Imbalanced class labels
  - Need to extract features from multiple sources



Existing Trust Predictors (Unsupervised)

- Unsupervised methods
  - Trust propagation is a key technique
  - But it requires sufficient connections for each user



Social Theories Can Help

- Homophily is one of the important theories that explain why users are connected
  - Similar users are likely to establish trust relations

- Exploiting homophily effect provides a fresh perspective for a novel trust predictor

- We first question if we can observe homophily in trust relations

Data Sets

- Two product review data sets
  - Epinions
  - Ciao

Table 1: Statistics of the Datasets

	Epinions	Ciao
# of Users	8,527	6,262
# of items	26,552	20,416
# of Ratings	225,579	167,320
# of Trust Relations	302,177	109,524
Max # of Trustors	1,285	100
Max # of Trustees	1,805	797
Trust Network Density	0.0042	0.0028
Clustering Coefficient	0.2242	0.2254

Assessing Existence of Homophily

- Are users with trust relations more similar in terms of their ratings than those without?

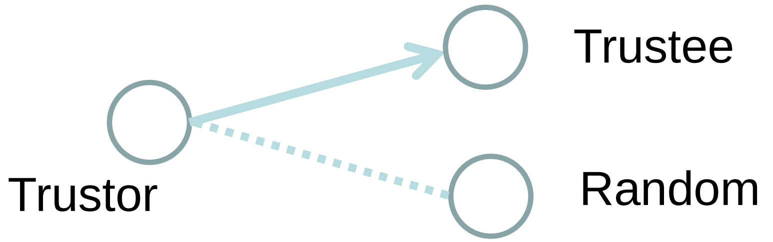
- Are users with higher similarity more likely to establish trust relations than those with lower similarity?

Findings for the First Question

- For each trust relation, we calculate two similarities

  - Similarity1: trustor and trustee

  - Similarity2: trustor and a random user



- We define two vectors **s** = { similarity1} and **t** = {similarity2}

- We conduct a one-tail t-test on **s** and **t**

$$H_0: \mathbf{s} = t; \qquad H_1: \mathbf{s} > \mathbf{t}$$

The null hypothesis is rejected at significance level 0.01 with p-value of 5.12e−18 and 3.76e−21 in Epinions and Ciao, respectively

DMML

ASU Arizona State University

Before we applying Homophily …

- We review **properties** associated with **trust**
  - – Correlation with user preference
  - – Transitivity, composability, and asymmetry
  - – Multiple facets (mTrust, WSDM2012)
  - – Evolution (eTrust, KDD2012)
- Next we propose a new model of trust

Modeling Trust

- Given that $\boldsymbol{u}_i$ is a k dimensional preference vector of u_i , the trust relation from u_i to u_j is modeled as the correlation of user preferences,

$$\mathrm{G}_{ij} \approx \mathrm{u}_i^T \mathrm{H} \mathrm{u}_j$$

This model can capture the following properties of trust,
- Correlation with user preference
- Transitivity, composability, and asymmetry
- Multiple facets

A Computational Interpretation of Homophily

- We need a bridge between homophily theory and our computational model.

- We define $\zeta(i, j)$ as homophily coefficient between u_i and u_j

  - $\zeta(i, j) \in [0,1], \ \zeta(i, j) = \zeta(j, i)$

  - The larger value $\zeta(i, j)$ is, the more likely a trust relation is established between u_i and u_j

Model Homophily

- We define homophily regularization to exploit homophily effect as

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta(i,j) \| \mathbf{u}_i - \mathbf{u}_j \|_2^2$$

A large homophily coefficient indicates that they are more likely to establish trust relations thus their preferences should be similar

Significance of Homophily Regularization

- Incorporating side information via homophily coefficient

  - User profile (gender, location, age, education)

  - Rating information (in our paper)

- Applying to cold-start users

  - For users with few or no trust relations, we still can get an approximate estimate of their preference via homophily regularization

$$\sum_{j=1}^{n} \zeta(i, j) \| \mathbf{u}_i - \mathbf{u}_j \|_2^2$$

Our Framework - hTrust

- hTrust is to solve the following problem

$$\min_{\mathrm{U},H} \| \mathrm{G} - \mathrm{U}^{\mathrm{T}}\mathrm{H}\mathrm{U} \|_F^2 + \alpha \| \mathrm{U} \|_F^2 + \beta \| \mathrm{H} \|_F^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} \zeta(i,j) \| \mathrm{u}_i - \mathrm{u}_j \|_2^2$$

$$s.t. \quad \mathrm{U} \geq 0, \ \mathrm{H} \geq 0$$

The first term captures the properties of trust
The fourth term is used to exploit homophily effect

After learning **U** and **H**, the likelihood of the *i*-th user trusts *j*-th user will be given by

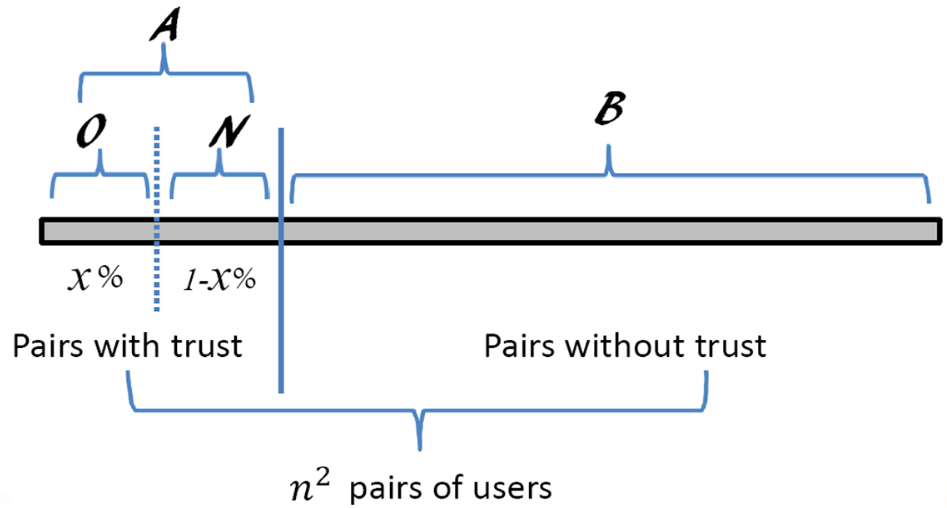
$$\tilde{\mathrm{G}}_{ij} = \mathrm{u}_i^T \mathrm{H} \mathrm{u}_j$$

How to Evaluate hTrust

- Is exploiting homophily effect helpful for trust prediction ?
- How does homophily regularization affect hTrust?
- How does homophily coefficient affect hTrust?

Experiment Setup

- Splitting data
  - x% as old trust relations
  - 1 - x% as new trust relations



- Evaluating steps
  - Ranking pairs of users in *N* and *B*
  - Choosing top-|*N*| ranked pairs as *C*
  - Calculating accuracy as

$$PA = \frac{|N \cap C|}{|N|}$$

Evaluating hTrust

- Is homophily effect helpful for trust prediction?
  - Comparing hTrust with the representative trust predictors
- How does homophily regularization affect hTrust?
- How does homophily coefficient affect hTrust?

Comparison of Different Trust Predictors

	50%	55%	60%	65%	70%	80%	90%
TP	0.1852	0.1897	0.1897	0.1845	0.1790	0.1663	0.1558
RS	0.1319	0.1230	0.1110	0.1029	0.0869	0.0813	0.0598
PCC	0.1160	0.1019	0.0884	0.0811	0.0614	0.0610	0.0469
JC	0.0940	0.0858	0.0725	0.0637	0.0480	0.0336	0.0279
simTP	0.2076	0.2105	0.2057	0.2011	0.1982	0.1857	0.1702
MF	0.2145	0.2121	0.2102	0.2057	0.1944	0.1837	0.1688
triNMF	0.2142	0.2129	0.2134	0.2064	0.1958	0.1875	0.1692
hTrust	0.2569	0.2517	0.2434	0.2326	0.2268	0.2072	0.1900
Random	0.0027	0.0026	0.0025	0.0024	0.0017	0.0015	0.0016

Comparison of Different Trust Predictors

	50%	55%	60%	65%	70%	80%	90%
TP	0.1852	0.1897	0.1897	0.1845	0.1790	0.1663	0.1558
RS	0.1319	0.1230	0.1110	0.1029	0.0869	0.0813	0.0598
PCC	0.1160	0.1019	0.0884	0.0811	0.0614	0.0610	0.0469
JC	0.0940	0.0858	0.0725	0.0637	0.0480	0.0336	0.0279
simTP	0.2076	0.2105	0.2057	0.2011	0.1982	0.1857	0.1702
MF	0.2145	0.2121	0.2102	0.2057	0.1944	0.1837	0.1688
triNMF	0.2142	0.2129	0.2134	0.2064	0.1958	0.1875	0.1692
hTrust	0.2569	0.2517	0.2434	0.2326	0.2268	0.2072	0.1900
Random	0.0027	0.0026	0.0025	0.0024	0.0017	0.0015	0.0016

Comparison of Different Trust Predictors

	50%	55%	60%	65%	70%	80%	90%
TP	0.1852	0.1897	0.1897	0.1845	0.1790	0.1663	0.1558
RS	0.1319	0.1230	0.1110	0.1029	0.0869	0.0813	0.0598
PCC	0.1160	0.1019	0.0884	0.0811	0.0614	0.0610	0.0469
JC	0.0940	0.0858	0.0725	0.0637	0.0480	0.0336	0.0279
simTP	0.2076	0.2105	0.2057	0.2011	0.1982	0.1857	0.1702
MF	0.2145	0.2121	0.2102	0.2057	0.1944	0.1837	0.1688
triNMF	0.2142	0.2129	0.2134	0.2064	0.1958	0.1875	0.1692
hTrust	0.2569	0.2517	0.2434	0.2326	0.2268	0.2072	0.1900
Random	0.0027	0.0026	0.0025	0.0024	0.0017	0.0015	0.0016

Questions

- Is exploiting homophily effect helpful for trust prediction ?

- How does homophily regularization affect hTrust?

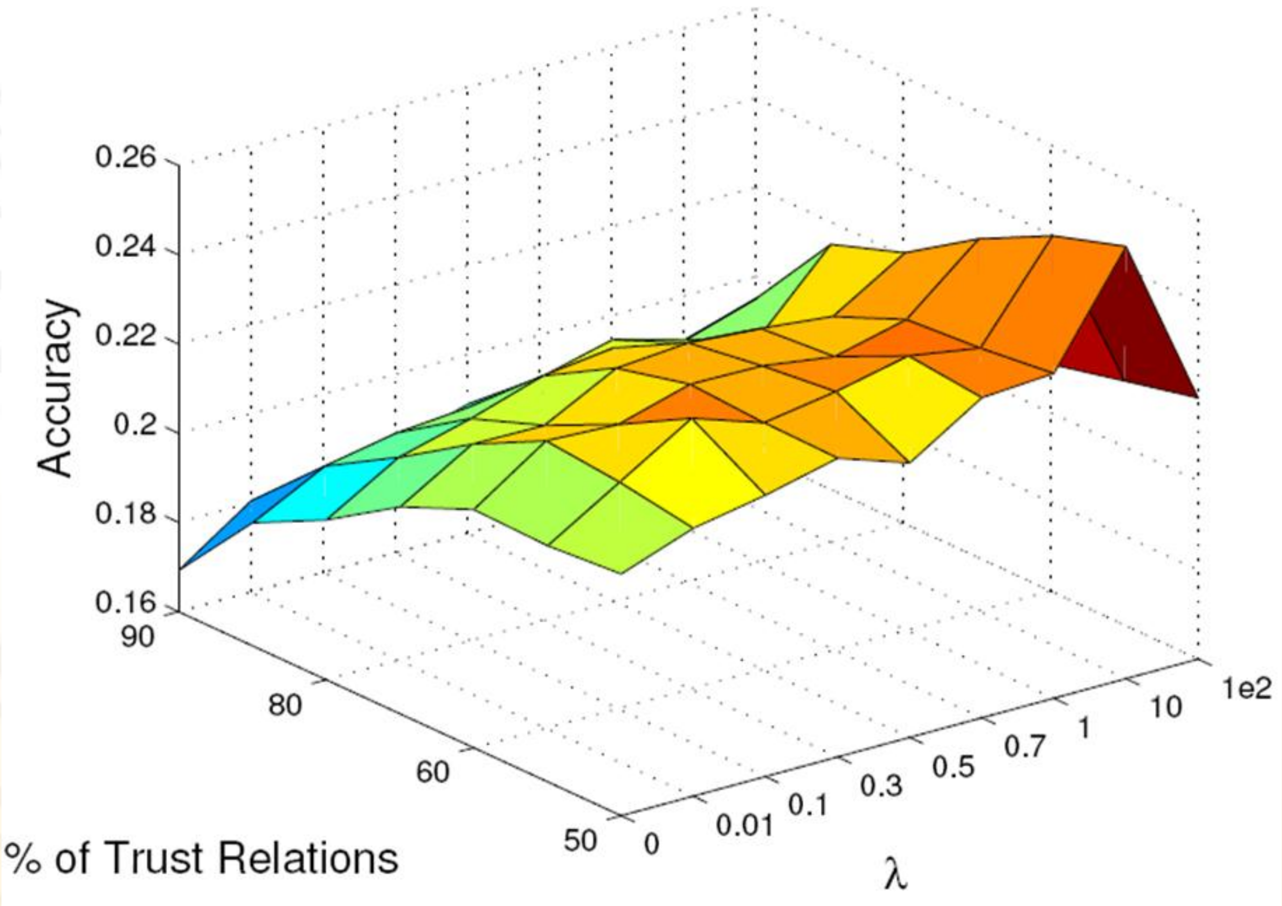
  - λ controls the contribution from homophily regularization thus we investigate the impact of λ on hTrust.

- How does homophily coefficient affect hTrust?

DMML

Arizona State University

Impact of Homophily Regularization



Questions

- Is exploiting homophily effect helpful for trust prediction ?

- How does homophily regularization affect hTrust?

- How does homophily coefficient affect hTrust?

  - Investigate different ways to obtain homophily coefficient.

Impact of Homophily Coefficient

Table 3: Different Measures of Homophily Coefficient. Note that $\zeta(i,j) = random$ means we randomly assign homophily coefficients ,while $\zeta(i,j) = 1$ indicates that homophily coefficients for all pairs of users are set to 1

Datasets		$\zeta(i,j) = JC(i,j)$	$\zeta(i,j) = PCC(i,j)$	$\zeta(i,j) = RS(i,j)$	$\zeta(i,j) = random$	$\zeta(i,j) = 1$
Epinions	50%	0.2382	0.2415	0.2569	0.2172	0.2192
	55%	0.2301	0.2354	0.2517	0.2153	0.2208
	60%	0.2227	0.2285	0.2434	0.2027	0.2071
	65%	0.2131	0.2196	0.2326	0.1907	0.1966
	70%	0.2019	0.2073	0.2268	0.1799	0.1856
	80%	0.1871	0.1937	0.2072	0.1558	0.1697
	90%	0.1732	0.1753	0.1900	0.1433	0.1498
Ciao	50%	0.1967	0.2098	0.2220	0.1630	0.1742
	55%	0.1941	0.2041	0.2193	0.1728	0.1721
	60%	0.1865	0.2069	0.2158	0.1585	0.1627
	65%	0.1780	0.1958	0.2082	0.1591	0.1613
	70%	0.1639	0.1820	0.1966	0.1479	0.1491
	80%	0.1441	0.1618	0.1749	0.1242	0.1304
	90%	0.1319	0.1502	0.1650	0.1214	0.1268

Impact of Homophily Coefficient

Table 3: Different Measures of Homophily Coefficient. Note that $\zeta(i,j) = random$ means we randomly assign homophily coefficients ,while $\zeta(i,j) = 1$ indicates that homophily coefficients for all pairs of users are set to 1

Datasets		$\zeta(i,j) = JC(i,j)$	$\zeta(i,j) = PCC(i,j)$	$\zeta(i,j) = RS(i,j)$	$\zeta(i,j) = random$	$\zeta(i,j) = 1$
Epinions	50%	0.2382	0.2415	0.2569	0.2172	0.2192
	55%	0.2301	0.2354	0.2517	0.2153	0.2208
	60%	0.2227	0.2285	0.2434	0.2027	0.2071
	65%	0.2131	0.2196	0.2326	0.1907	0.1966
	70%	0.2019	0.2073	0.2268	0.1799	0.1856
	80%	0.1871	0.1937	0.2072	0.1558	0.1697
	90%	0.1732	0.1753	0.1900	0.1433	0.1498
Ciao	50%	0.1967	0.2098	0.2220	0.1630	0.1742
	55%	0.1941	0.2041	0.2193	0.1728	0.1721
	60%	0.1865	0.2069	0.2158	0.1585	0.1627
	65%	0.1780	0.1958	0.2082	0.1591	0.1613
	70%	0.1639	0.1820	0.1966	0.1479	0.1491
	80%	0.1441	0.1618	0.1749	0.1242	0.1304
	90%	0.1319	0.1502	0.1650	0.1214	0.1268

Our Contributions

- Verifying the existence of homophily in trust relations
- Providing a principled way to incorporate homophily effect into a computational model and proposing hTrust
- Evaluating hTrust to understanding the work of hTrust

Future Work

- Applying homophily regularization to supervised trust predictors
- Exploiting homophily effect for other kinds of relations
  - Following relations in Twitter
  - Friendship in Facebook
- Investigating the dynamics of user preferences
  - Temporal information

Questions

Acknowledgments: This work is, in part, sponsored by ARO (#025071) and NSF (#IIS-1217466). Comments and suggestions from DMML members and reviewers are greatly appreciated.