Addressing Task Heterogeneity in Social Media Analytics

Arun Reddy Nelakurthi, Jingrui He

Arizona State University

anelakur@asu.edu, jingrui.he@asu.edu

BACKGROUND

1. **Social media**

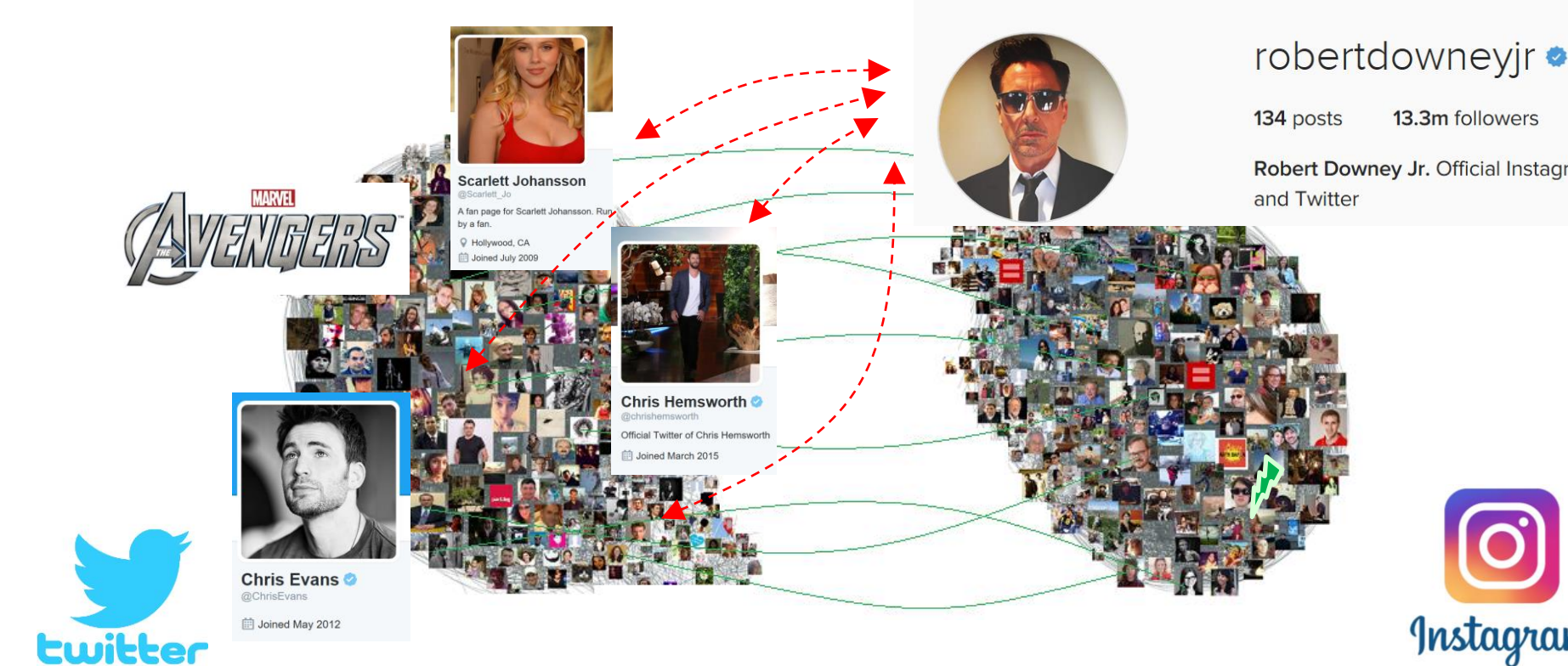2. **Diabetes dedicated social networks**
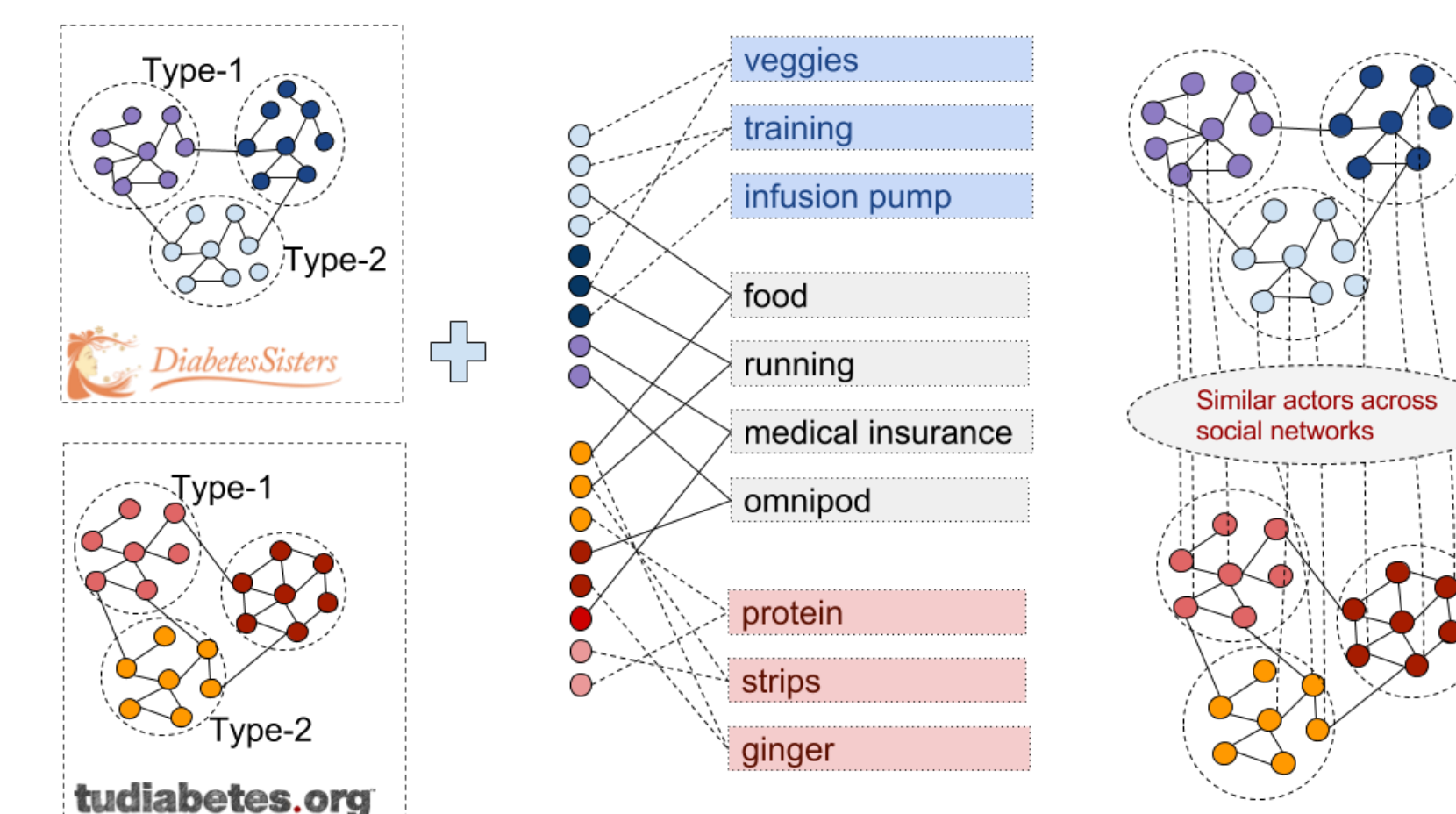

3. **User behavior modeling**

Optimistic — Pessimistic

4. **Impact of social media**

SIMILAR ACTOR RECOMMENDATION

1. **Setup**

2. **Cross-network links**

veggies / training / infusion pump / food / running / medical insurance / omnipod / protein / strips / ginger

Similar actors across social networks

tudiabetes.org

3. **Joint Matrix Tri-factorization**

$$\mathbf{X} \approx \mathbf{F} \, \mathbf{S} \, \mathbf{G}^T$$

Users / Keywords / User groups / Keyword groups

Reconstruction error   Graph regularizer

$$\min \sum_{k=1}^{K} \left\{ \left\| \mathbf{X}_k - \mathbf{F}_k \mathbf{S} \mathbf{G}^T \right\|_F^2 + \alpha_k tr\left(\mathbf{F}_k^T \mathbf{L}_k \mathbf{F}_k \right) \right\}$$

$$\text{s.t.} \quad \mathbf{F}_k \geq 0, \mathbf{S} \geq 0, \mathbf{G} \geq 0$$

$$\mathbf{F}_k^T \mathbf{D}_F \mathbf{F}_k = \mathbf{I}_F, \sum_j \mathbf{G}_{i,j} = 1, k = 1, \ldots, K.$$

Non-negative constraints
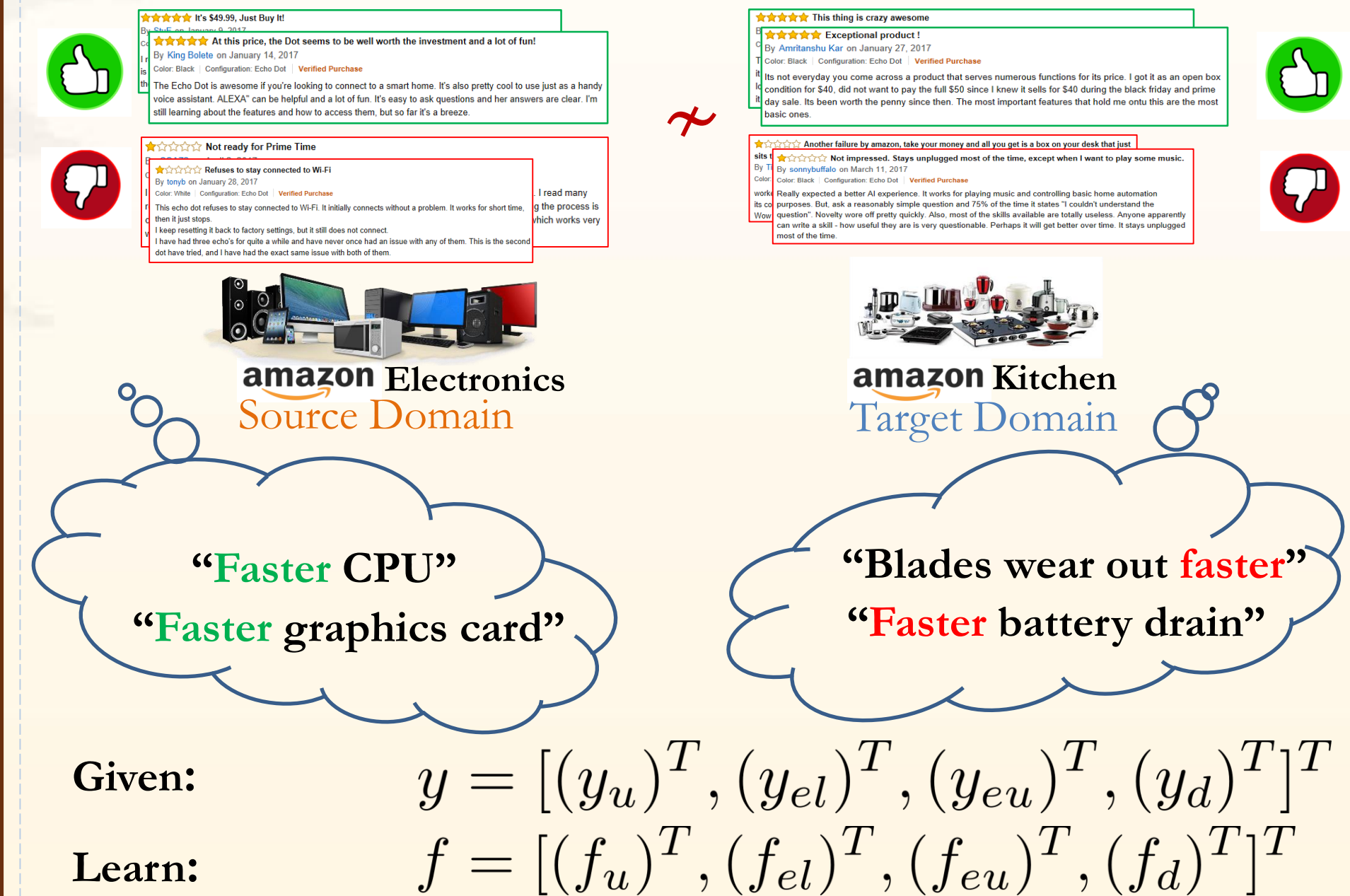
Orthogonality constraints

3. **Similar actor recommendation**

A is close to B   B and F are similar   A is close to F

Case-study on Diabetes-dedicated social networks

tudiabetes.org   DiabetesSisters
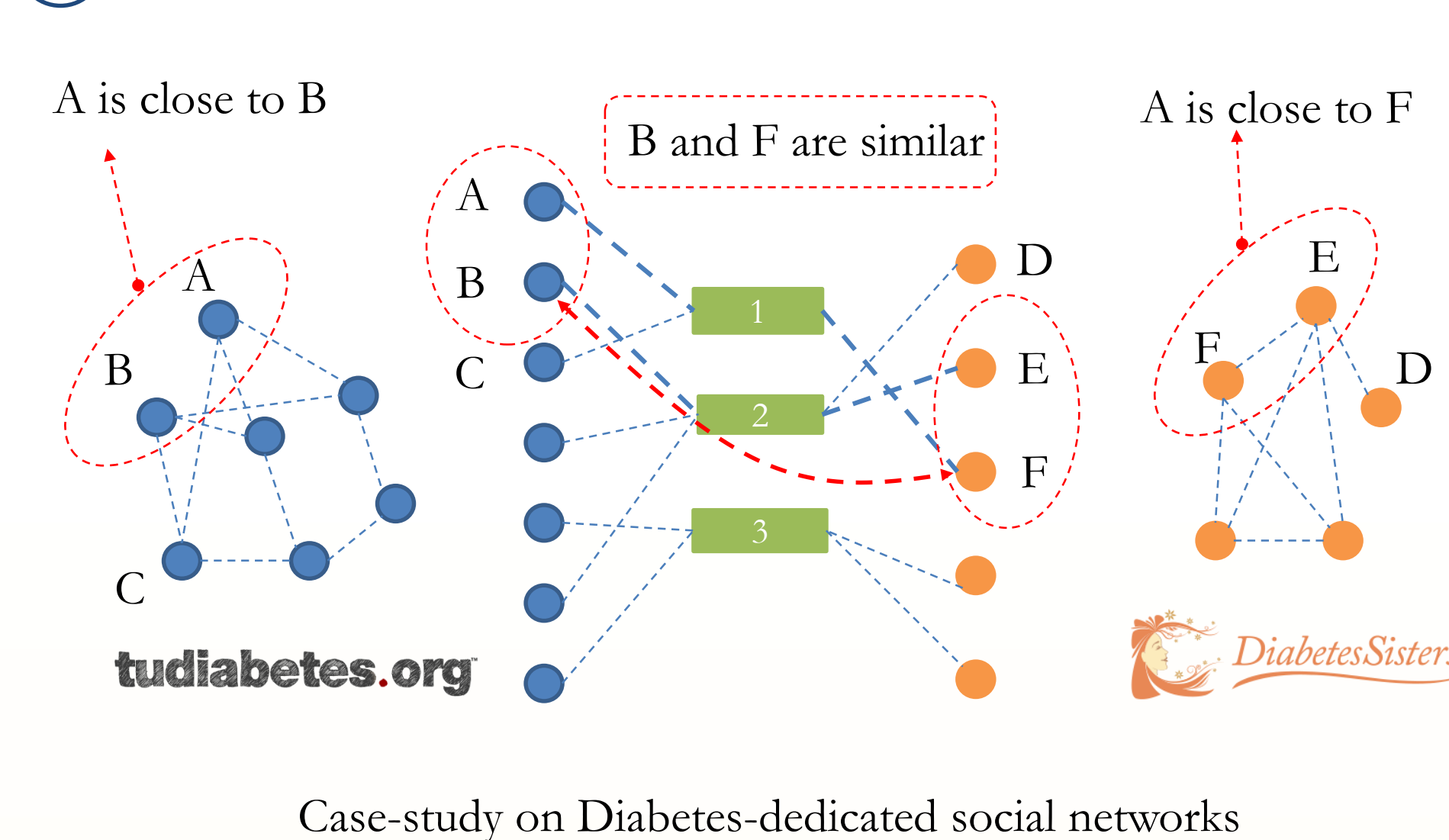
MODELING USER BEHAVIOR

1. **Setup**

amazon Electronics **Source Domain**   amazon Kitchen **Target Domain**

"**Faster** CPU"
"**Faster** graphics card"

"Blades wear out **faster**"
"**Faster** battery drain"

Given: $y = [(y_u)^T, (y_{el})^T, (y_{eu})^T, (y_d)^T]^T$

Learn: $f = [(f_u)^T, (f_{el})^T, (f_{eu})^T, (f_d)^T]^T$

2. **User-examples-keywords tripartite graph**

Source Domain $\mathcal{X}_S$

$\mathcal{U}^S$ / $\mathcal{U}_C$ / $\mathcal{U}^T$

Target Domain $\mathcal{X}_T$

Users   Examples   Keywords

3. **Objective function**

$$Q(f) = \frac{1}{2}\sum_i^e w_i \sum_j^u v_j \mathbf{A}_{i,j}^{(3)}\left(\frac{f_i}{\sqrt{\mathbf{D}_i^{(3)}}} - \frac{f_j}{\sqrt{\mathbf{D}_j^{(3)}}}\right)^2$$

Users - Examples

$$+ \frac{1}{2}\sum_i^e w_i \sum_j^d \mathbf{A}_{i,j}^{(3)}\left(\frac{f_i}{\sqrt{\mathbf{D}_i^{(3)}}} - \frac{f_j}{\sqrt{\mathbf{D}_j^{(3)}}}\right)^2$$

Examples - Keywords

$$+ \mu \sum_k^{u+e+d} (f_k - y_k)^2$$

Regularizer on known labels and priors

$$= f^T(I_{(u+e+d)\times(u+e+d)} - \mathbf{S}^{(3)})f + \mu \|f - y\|^2$$

4. **Instance reweighting to fix covariate shift**

$$w_i = \frac{P_{target}(\mathbf{x}_i)}{P_{source}(\mathbf{x}_i)}$$

5. **User consistency**

Higher weights to users with consistent labeling behavior

$$v = \sum_i^{el^S} \sum_j^{el^T} \text{sim}(\mathbf{x}_i, \mathbf{x}_j) \times y_i \times y_j$$

$\mathbf{x}_i \in X^S$
$\mathbf{x}_j \in X^T$
$v \in [0,1]$

IMPACT OF SOCIAL MEDIA

1. Web-based and in-clinic surveys on patients with Diabetes Mellitus(DM).

2. Questions included reasons for visiting the social media websites, and ascertainment of DM reported self-care practices.

3. 45 Online participants and 167 in-clinic participants.

Top reasons to use social media

(a) Offer support   (b) Seek support
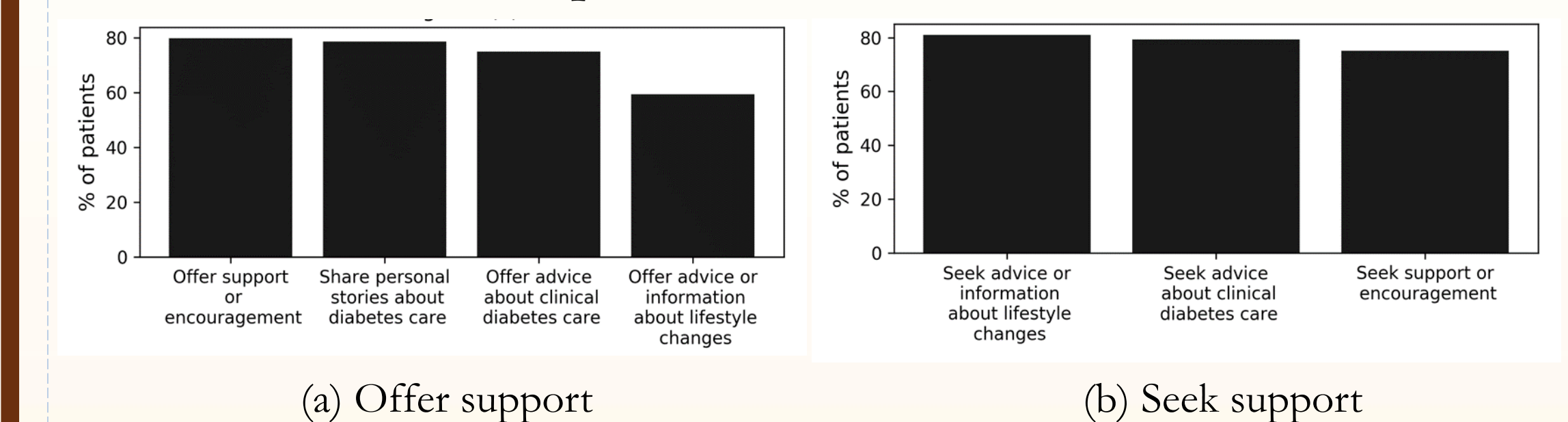
REFERENCES

Nelakurthi, Arun Reddy, et al. "User-guided Cross-domain Sentiment Classification." *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2017.

Nelakurthi, Arun Reddy, and Jingrui He. "Finding Cut from the Same Cloth: Cross Network Link Recommendation via Joint Matrix Factorization." *AAAI*. 2017.

Nelakurthi, Arun Reddy, et al. "Usage of Social Media for Diabetes Management in a Clinic-Based Population." *DIABETES*. Vol. 66. 1701 N BEAUREGARD ST, ALEXANDRIA, VA 22311-1717 USA: AMER DIABETES ASSOC, 2017.

Nelakurthi, Arun Reddy, et al. "Impact of Social Media on Behaviors of Patients with Diabetes." *DIABETES*. Vol. 65. 1701 N BEAUREGARD ST, ALEXANDRIA, VA 22311-1717 USA: AMER DIABETES ASSOC, 2016.