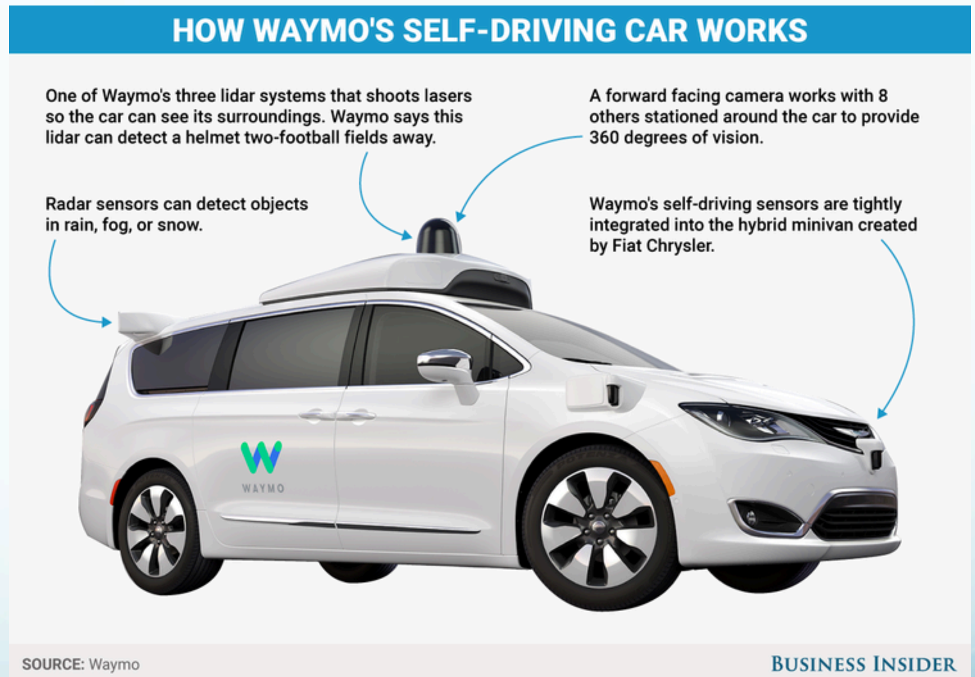
Some Adventures in Using Constraints in Machine Learning and Data Mining

Ian Davidson
UC Davis

Acknowledgements

- Thank you for the invitation.
- Lots of Smarter Collaborators
  - Research - University of Virginia (Ravi), Google Research (Basu), University de Orleans (Bich, Christel Vrain), University de Lyon (Julien Velcin), University of Louvain (Siegfried Nijssen, Pierre Schaus)
  - Applications – Naval Medical Research Center, UC Davis Center for Imaging, Pennington Institute for Neuroscience, SoarTech, Google Ads Team
- Funding
  - 2019- National Science Foundation – "Explaining Unsupervised Learning – Combinatorial Optimization Approaches"
  - 2019- Google – "Combining Symbolic Reasoning and Deep Learning: A Constrained Optimization Formulation"
  - 2018 - Office of Naval Research – "Deep Graph Learning"
  - In the past NSF, OSD, Google, Yahoo, ONR

Machine Learning Has Many Successes



And It Can Be Quite Easy

- My student Aubrey Gress spent **a summer** working in a group to get the Google car to read street signs.

Signs are designed to be easy to tell apart
Lots of examples
Easy to accurately label

But It Is Not All Roses – Consider Cognitive Network Discovery

fMRI = Movie of Your Brain

Science – What cognitive networks are associated with tasks

Selection – Job selection by network strength

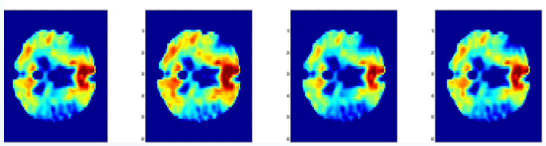
Diagnosis – Network biomarkers for Alzheimer’s

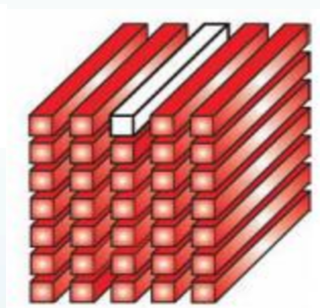
Treatment – Patterns amongst network · Precision medicine

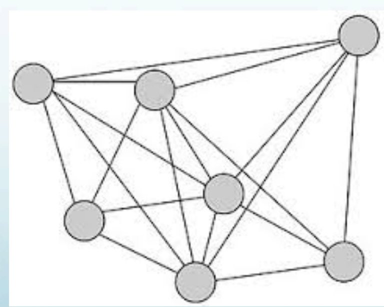
From fMRI Data to Graphs

[With NMRC, Pennington Institute, UC Davis Medical Imaging Group]

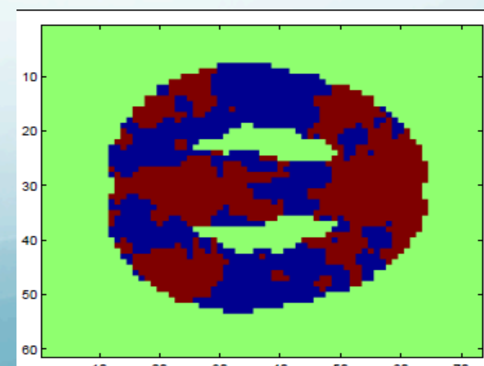
Can We Discover What Cognitive Networks Are Associated With Tasks?





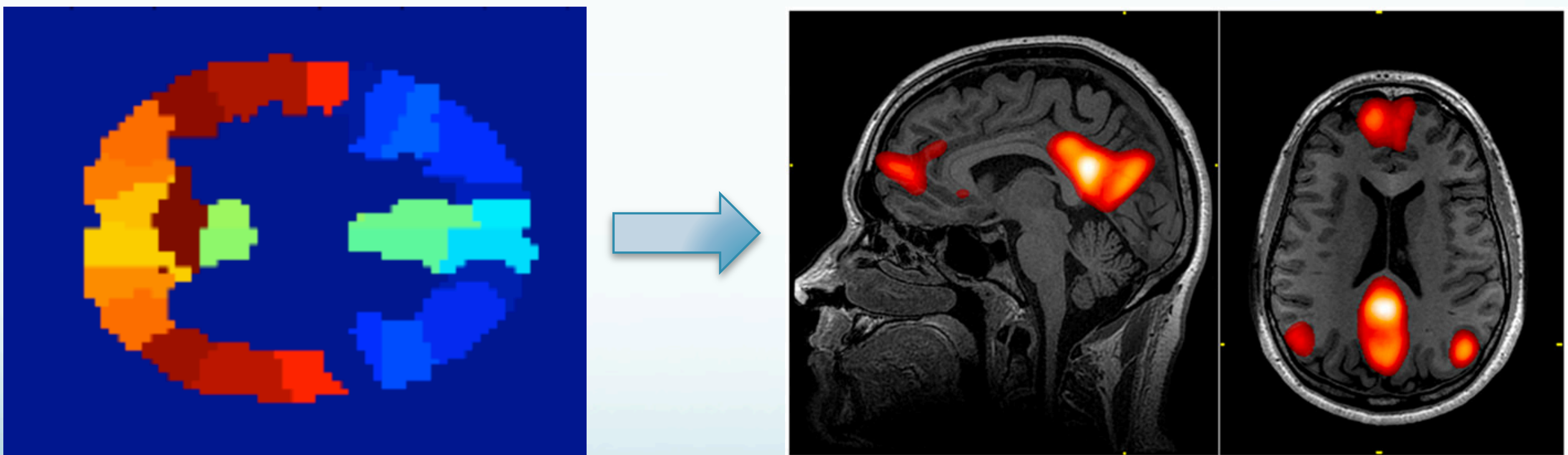


Network Activity vs Background



But There Are Some Challenges – #1 Adding Guidance

- Results must be **consistent** with biological knowledge (brain atlases)



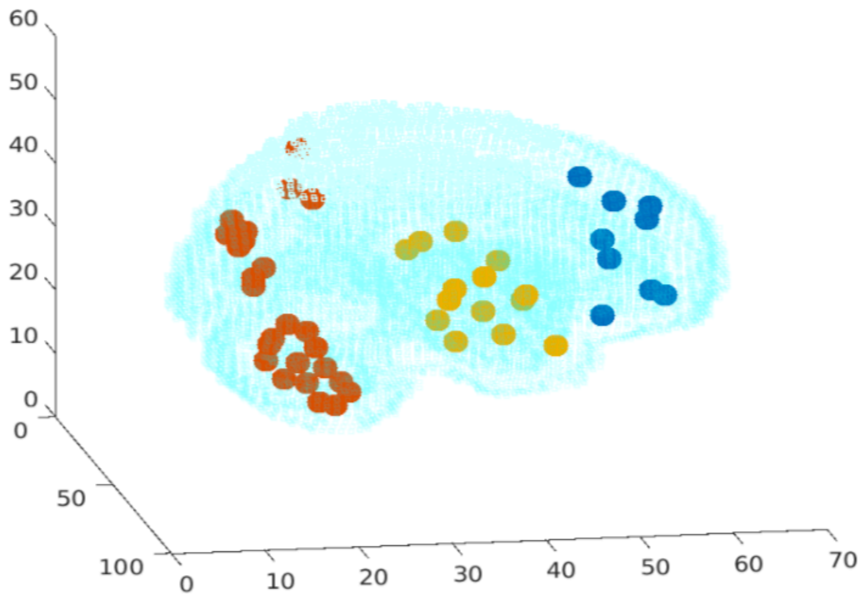
But There Are Some Challenges - # 2 Explanation

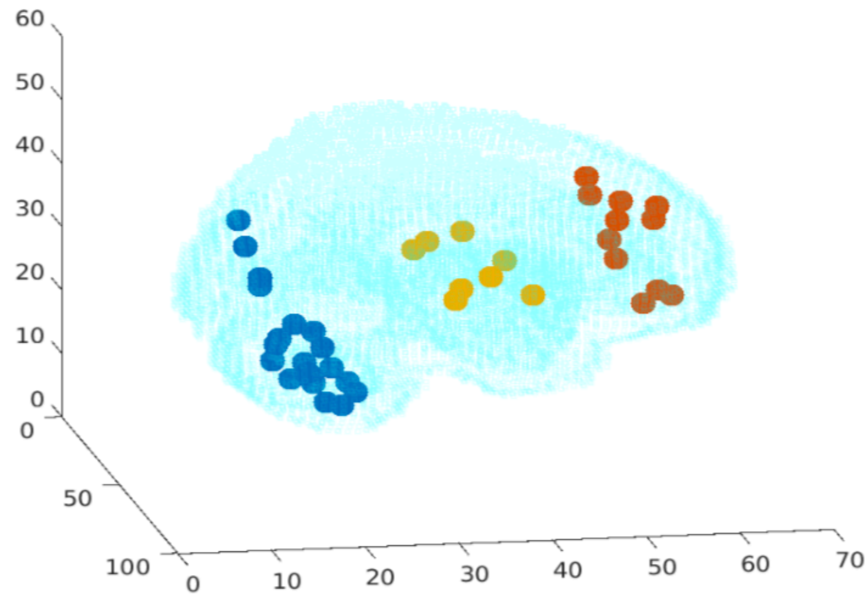
- Results must be **explainable**
- Precision medicine approach for SZ treatment
  - Explain that the results are **fair** to the FDA (or equivalent).
    - Are we giving expensive treatments disproportionately to some groups?
  - Detailed **explanation** to our collaborators to determine neurologically plausible.
    - Reduced connectivity between ROI x and y
  - High level explanations to build **trust** by other practitioners (i.e. best practices committee at HMO).
    - Dampening of connectivity in Executive functional network

Joint work with U.C. Davis Imaging Center

But There Are Some Challenges – #3 Human in the Loop Extensions

- Algorithm's output raises questions
  - Why did you find this?
  - Can you make these changes?



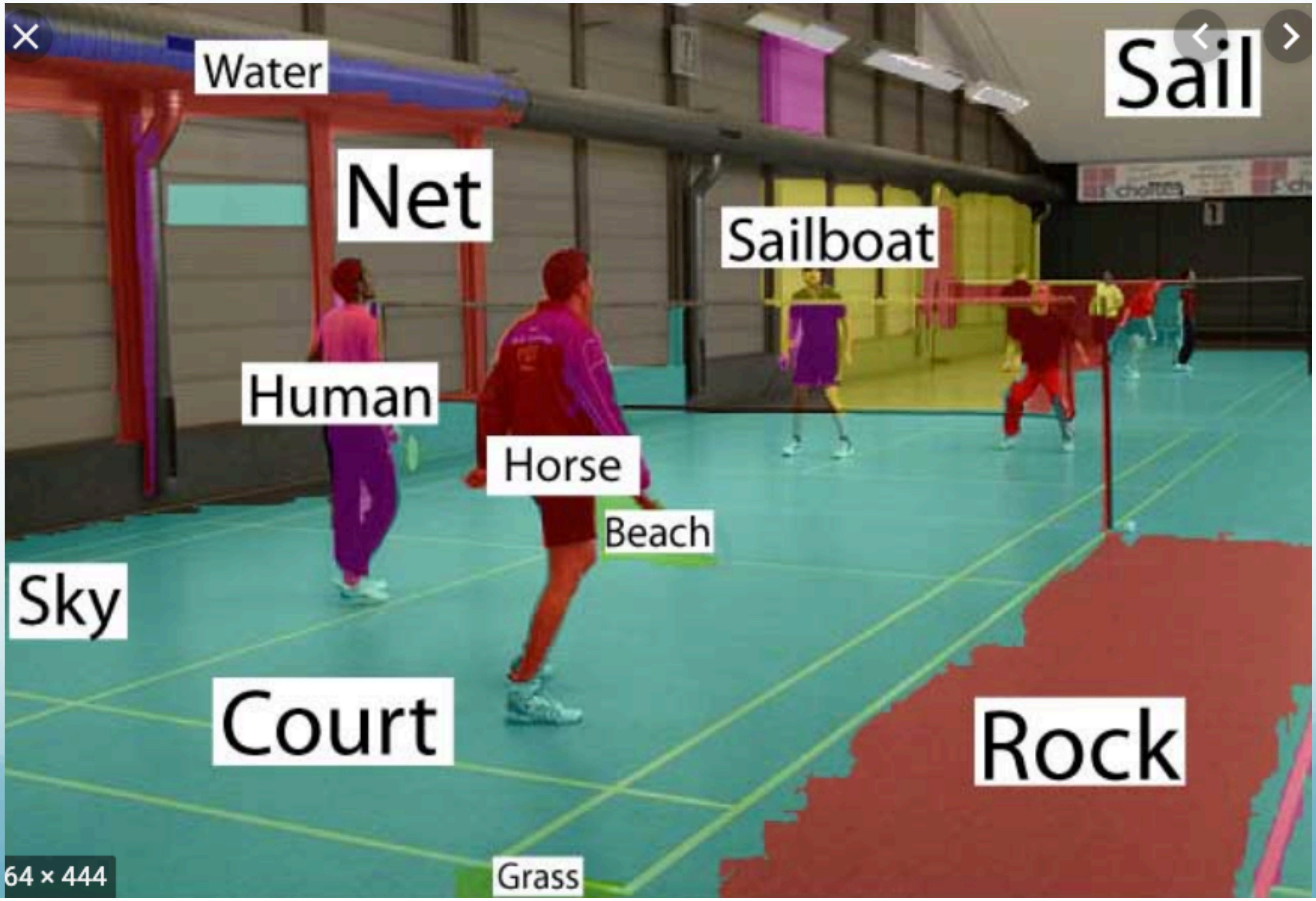


Found by Machine

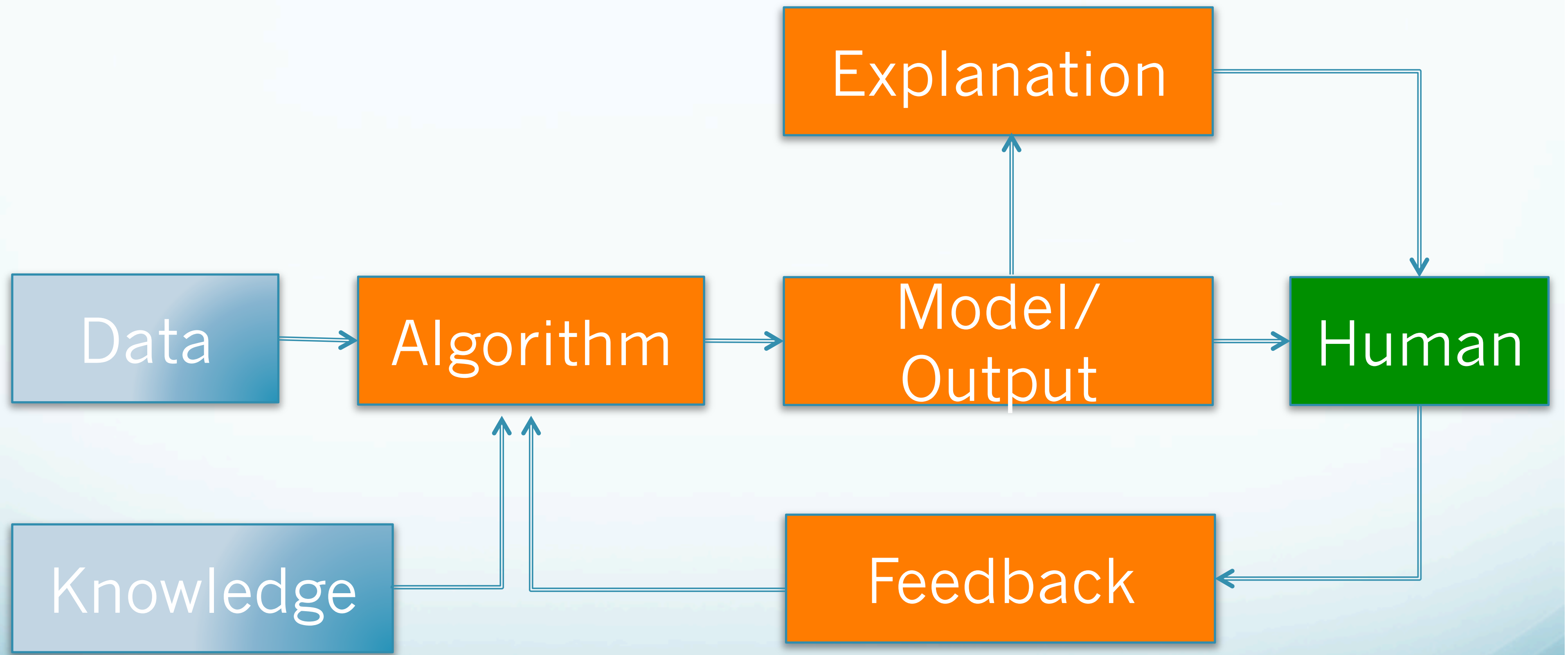
Suggested Refinement by Human

Other Examples Where Some of These Challenges Arise

Tagging with deep learning

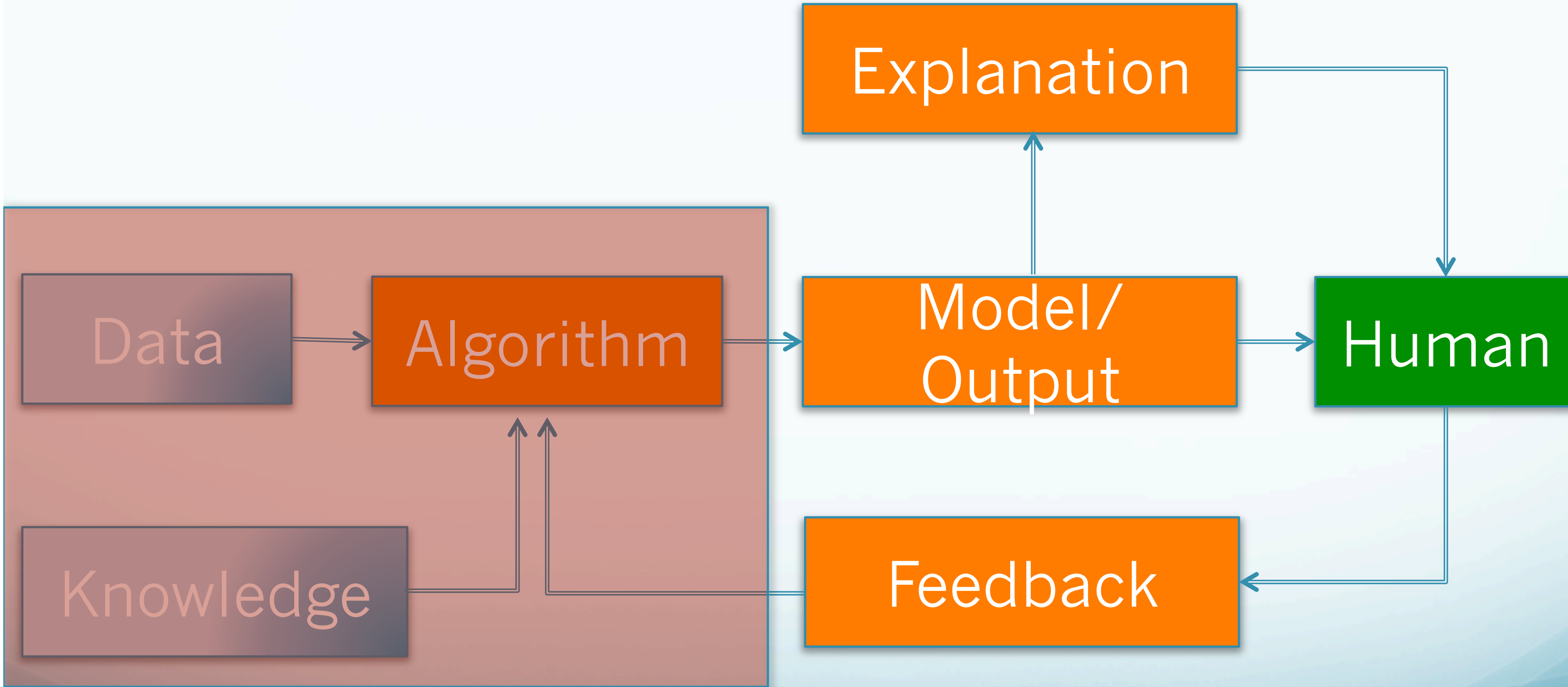


A Future of Machine Learning and Data Mining



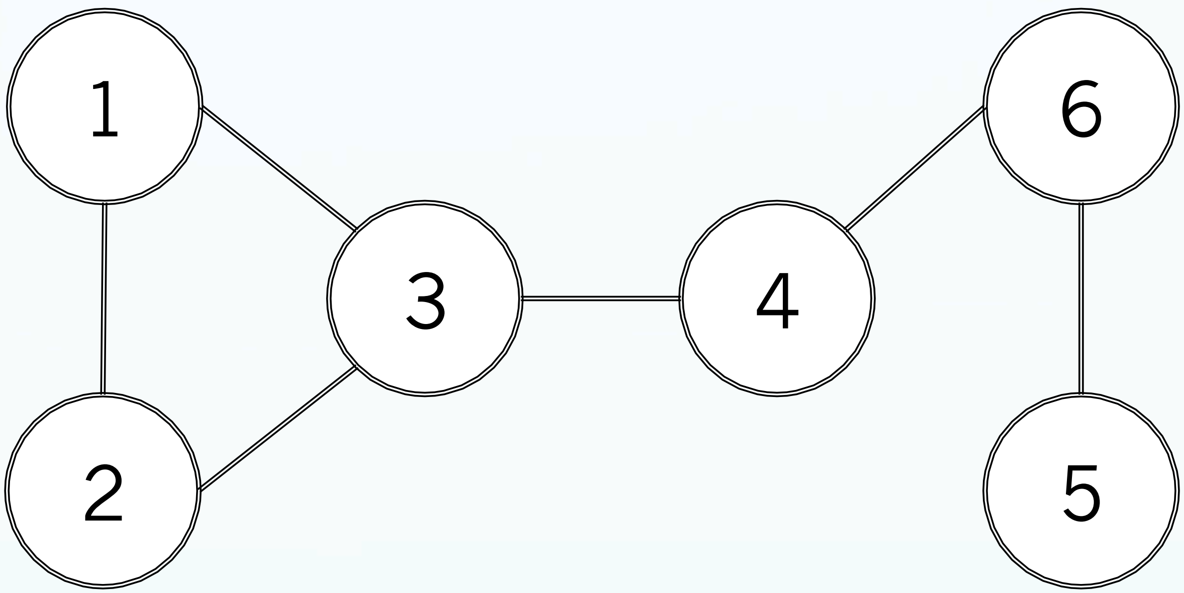
Adventures in Formulating These Challenges as Constrained Optimization Problems

Adventure #1 - Adding Semantic Knowledge as Constraints



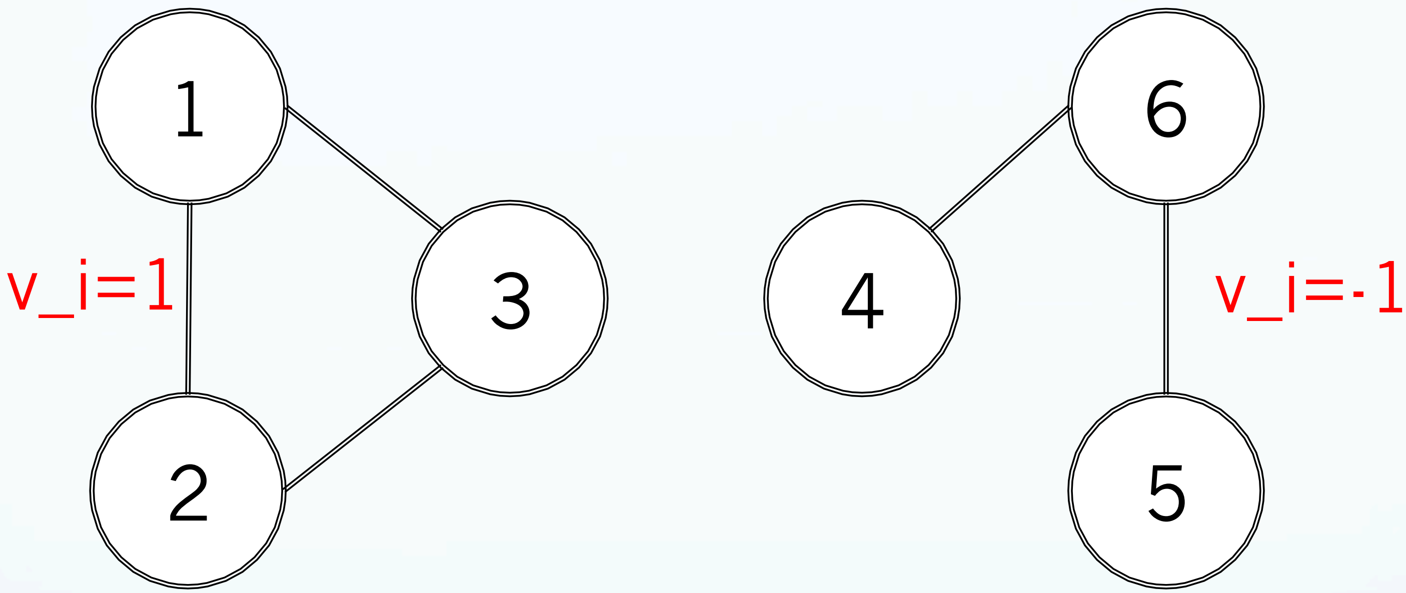
Adventures in Formulating These Challenges as Constrained Optimization Problems

A Simple Example of Constrained Clustering



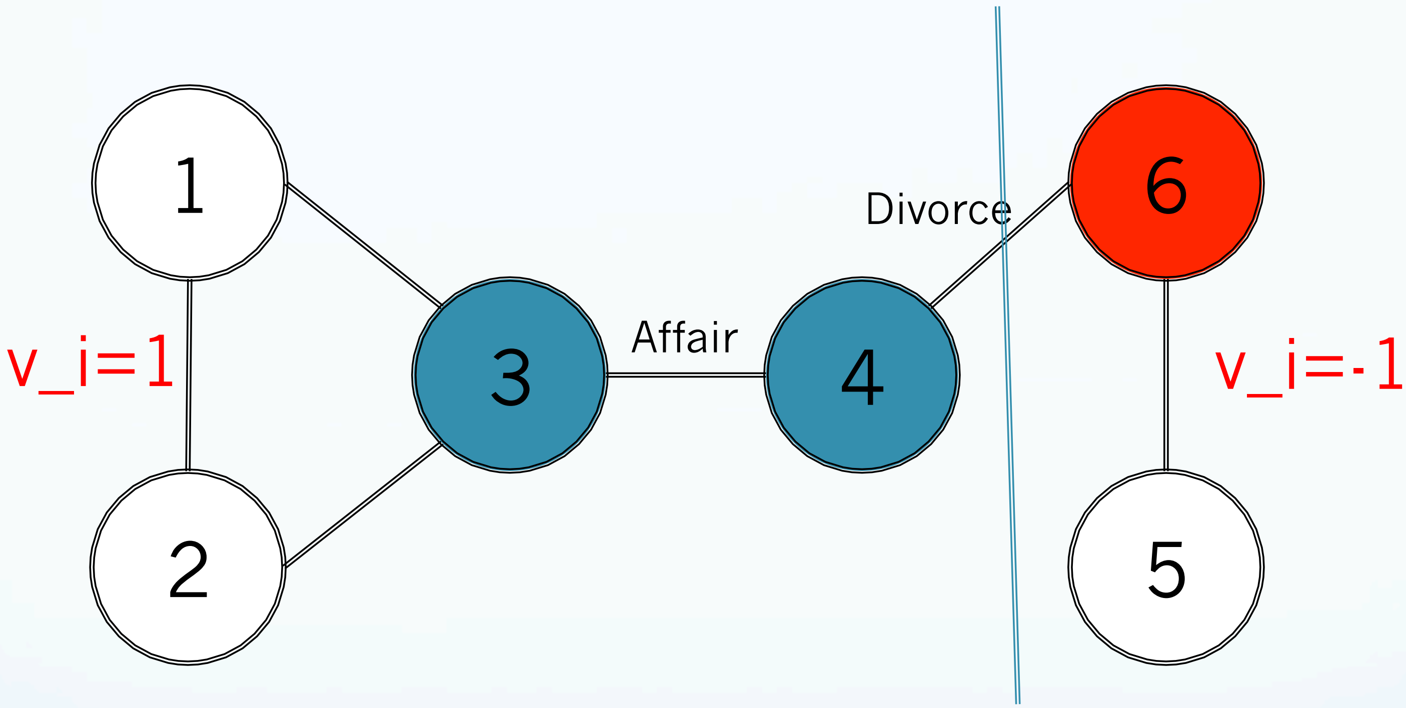
Imagine this is your ego network in Facebook. Want to Create two dinner parties (becomes your apartment is small)

A Simple Example of Constrained Clustering



Voilà! Your Two Dinner Party Guest List… v is an indicator vector

But It's Never That Easy



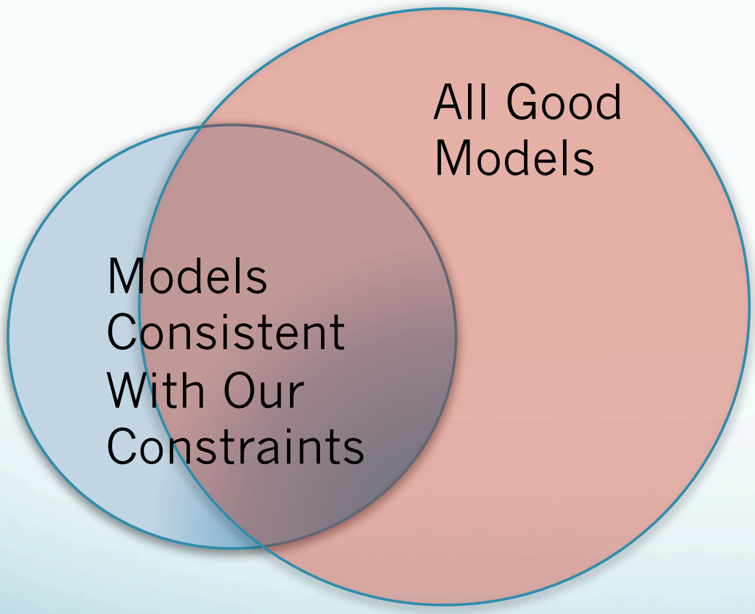
Maybe 6 divorced 4 because they were having an affair with 3

Find a *constrained cut* that:

- Minimizes the cost (fit to data)
- Satisfies these constraints (fit to human guidance)

Computationally What We Are Doing To Our Search

Restricting Search Space
But not hints from
optimal solution



Why?

- To enforce knowledge (human generated)
  - Satisfy requirements to make results useful/interesting
- Correct objective function issues (side information generated)
  - Objective functions in ML/DM are mathematically convenient
- Overcome lack of training data (transfer learning)
  - Use results from source domain to help learning in target domain

Lots of Previous Work

Input

Dataset of Instances Together/Apart Constraints

$x_1 \ldots x_n$

$T : List\ of\ pairs$

$A : List\ of\ pairs$

Variety of Solvers

Heuristics
K-Means: Wagstaff, Cardie ICML 2001
EM: Basu et' al. KDD 2007

Mathematical Programming
Spectral Clustering: Wang, Davidson, KDD 2010

Deep Learning
Zhang, Davidson ECML 2019

Output

Clustering That Satisfies All/Most/ Some of Constraints

$\Pi_1, \Pi_2 \ldots \Pi_k$

$\forall (i,j) \in T, C(i) = C(j)$

$\forall (i,j) \in A, C(i) \neq C(j)$

Three Generations of Enforcing Constraints

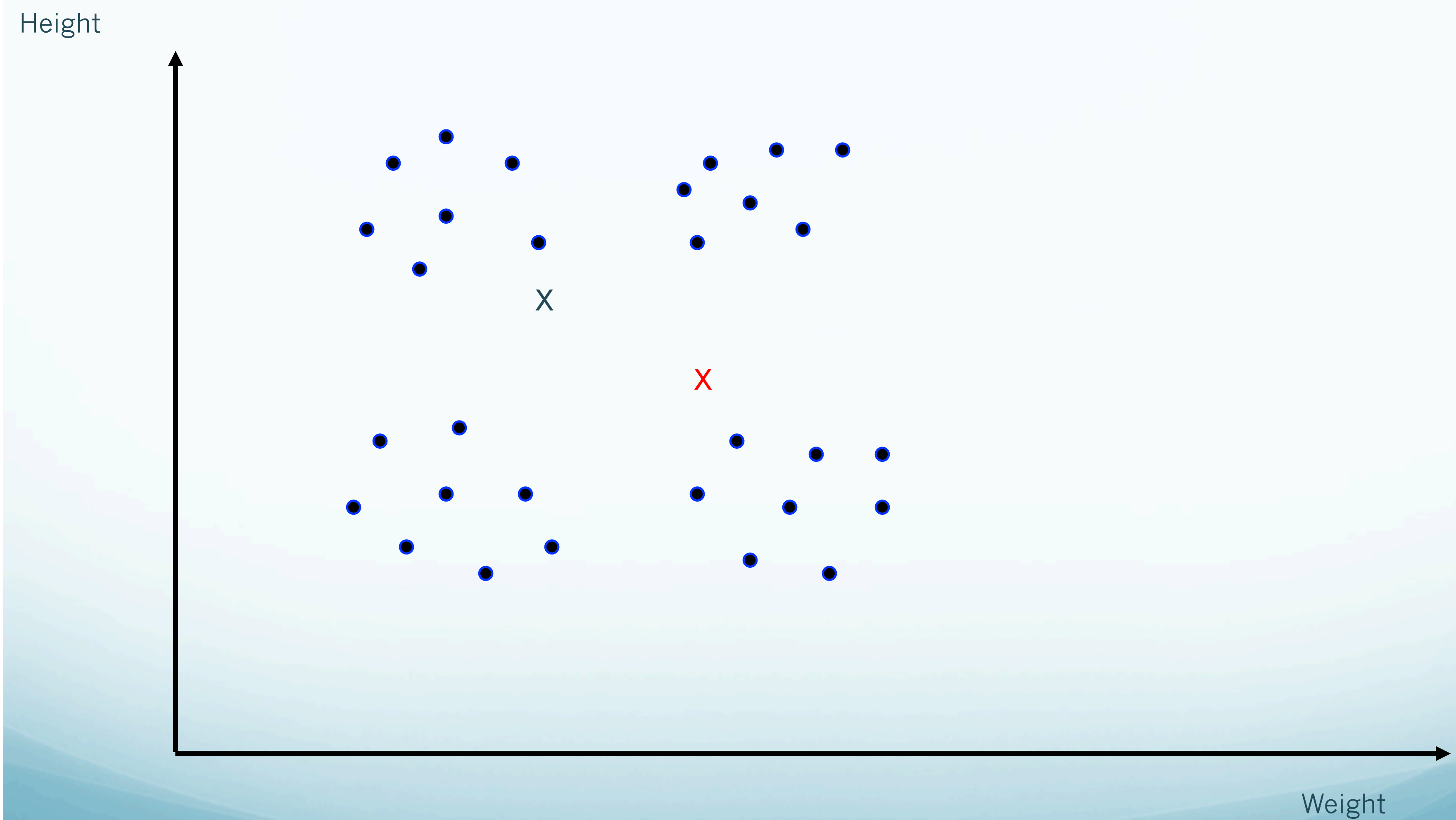
- Generation #1 – Existing algorithms
  - Adding constraint satisfaction step
  - Parititional - COP-kmeans [Wagstaff ICML 2000]
  - Hierarchical - Gilpin and Davidson [KDD 2011]
- Generation #2 – New Relaxed Formulations
  - New formulations
    - MP formulation after relaxing the problem
      - Wirth [ICML 08], Davidson [KDD 10], De Bie [JMLR 07]
- Generation #3 – Just starting
  - Deep learning formulation – complex constraints

K-Means Algorithm

1. Randomly choose a centroid for each cluster
2. Loop
    - Assign the instance to the closest centroid
    - Recalculate centroids
3. Goto 2 until distortion is small

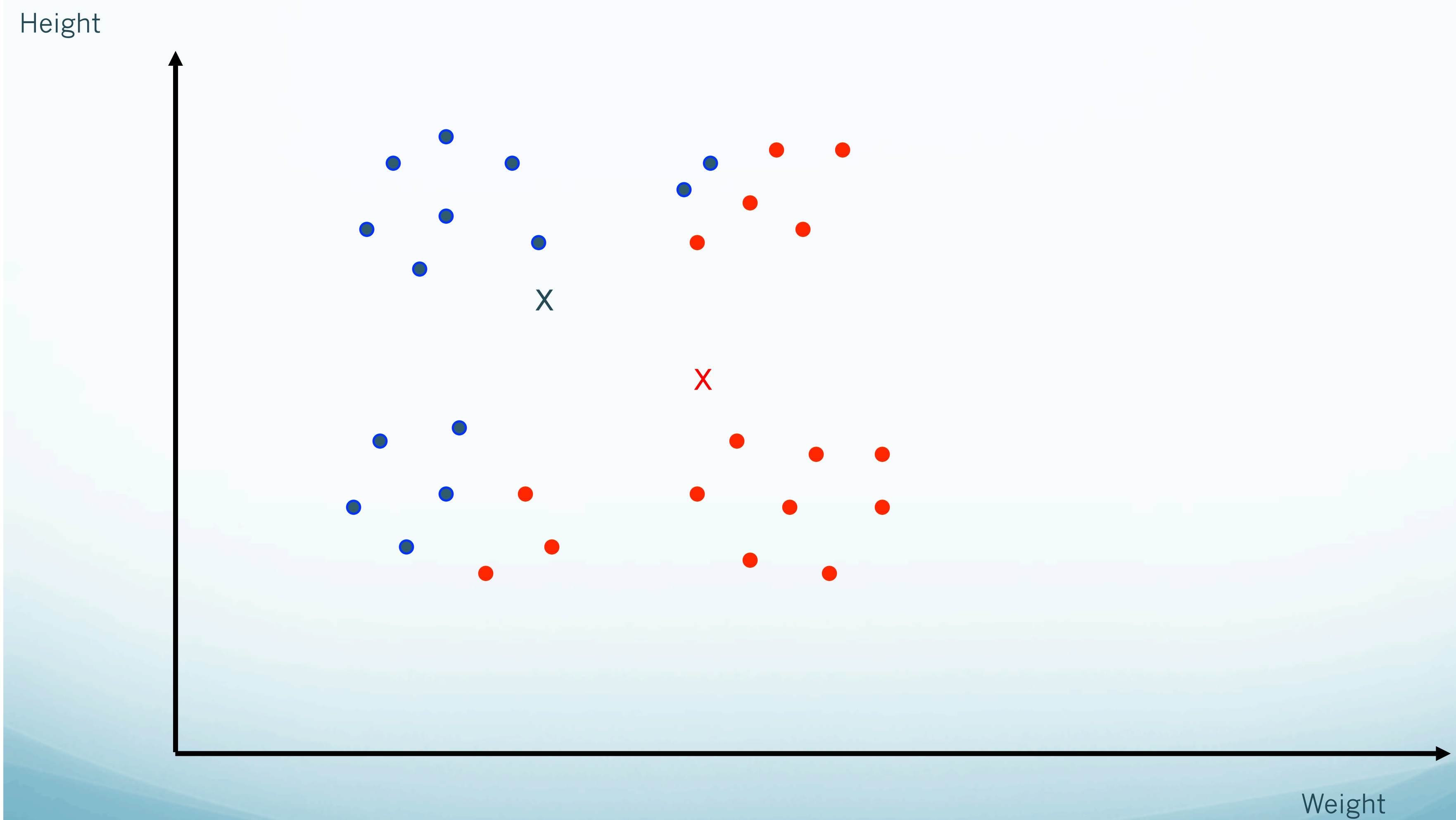
K Means Example (k=2)

Initialize Means

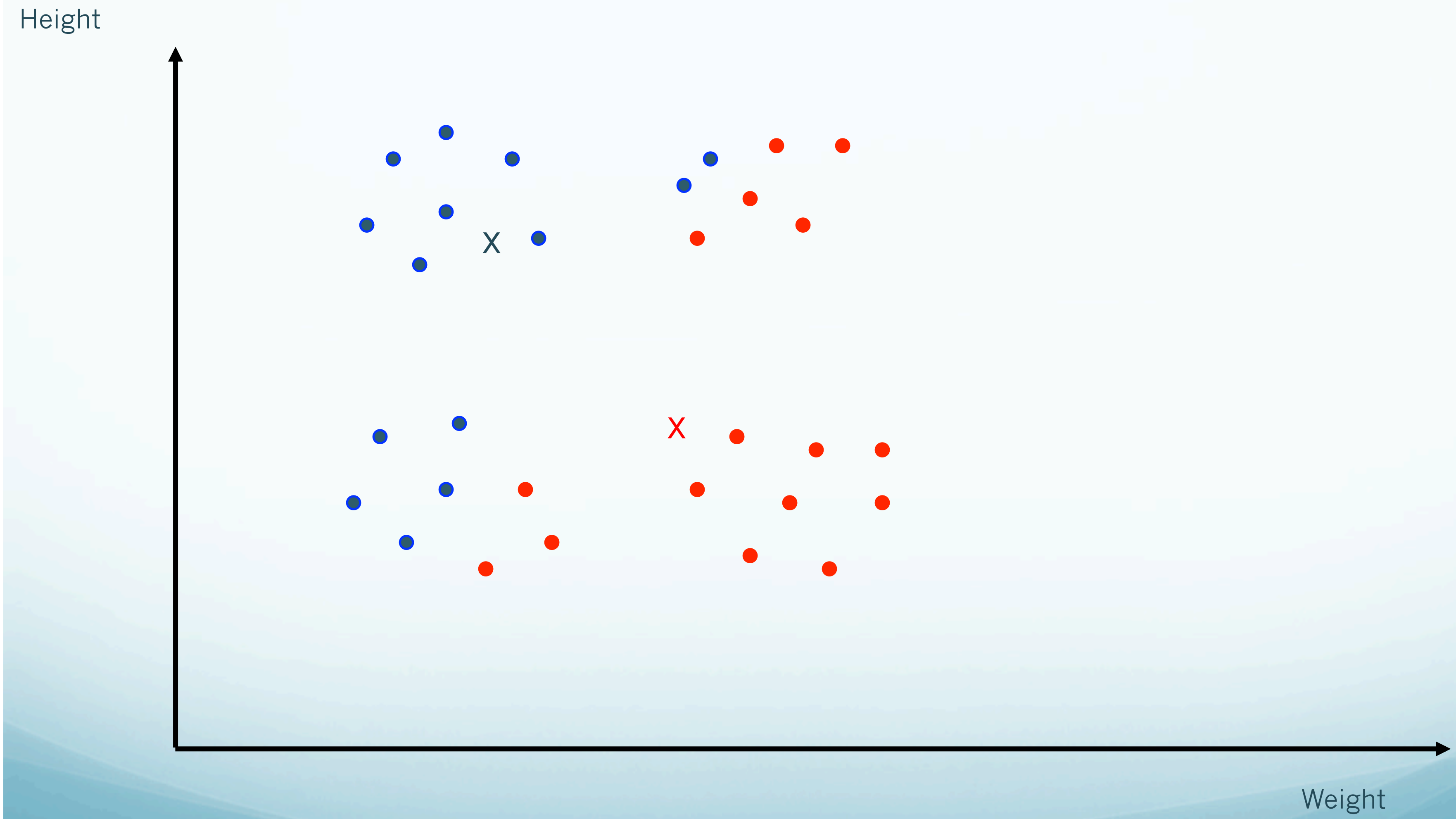


K Means Example

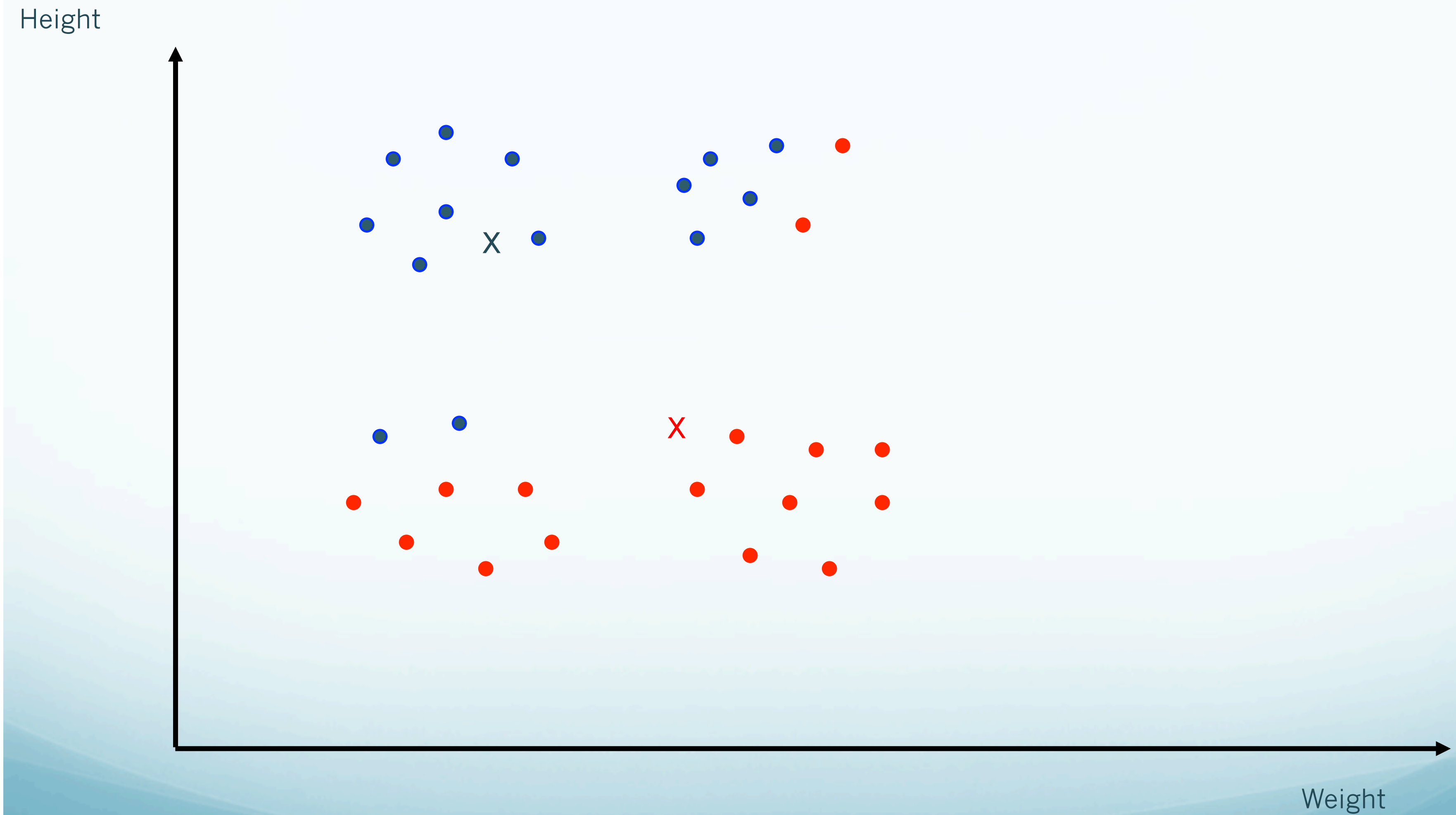
Assign Points to Clusters



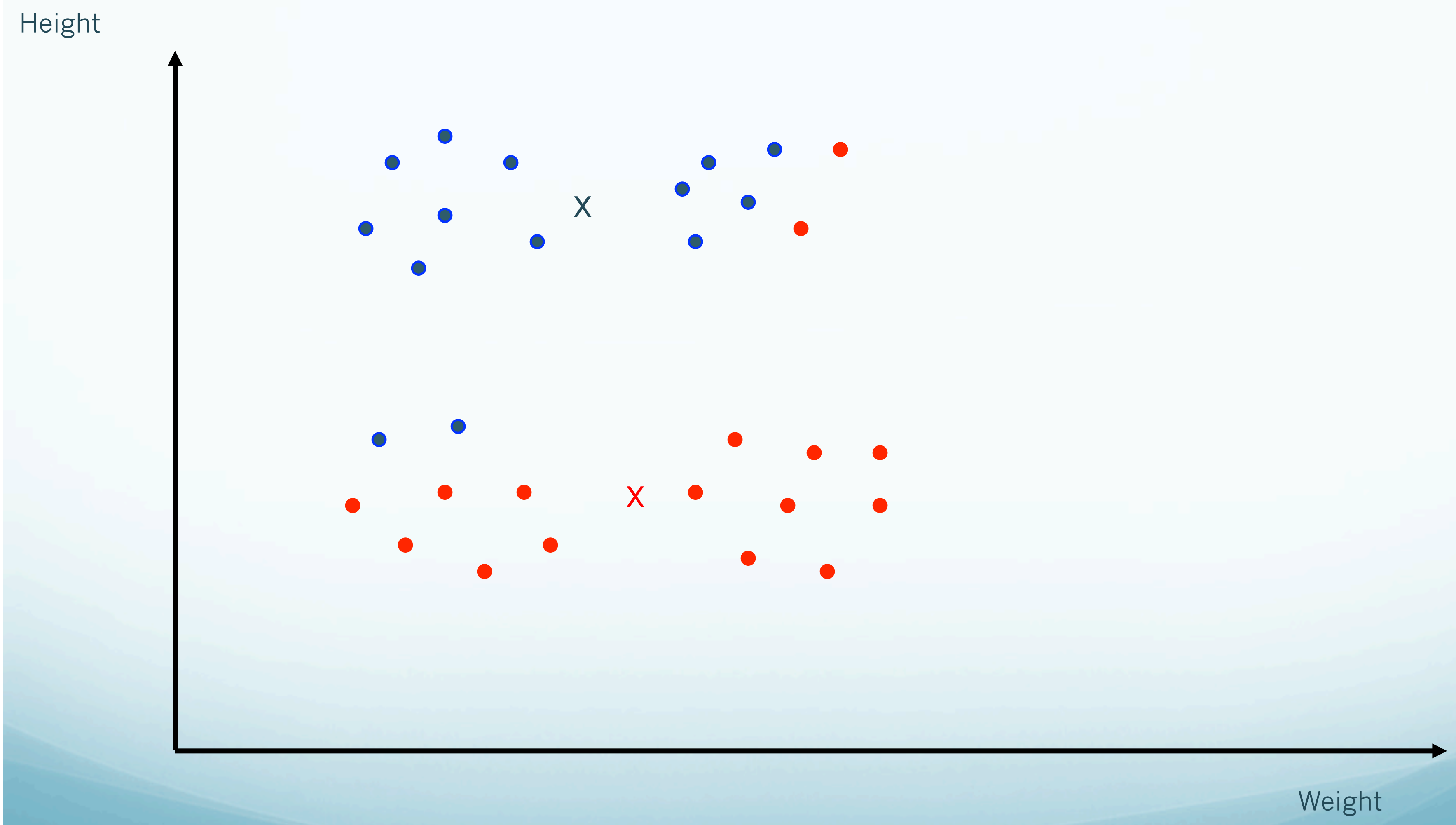
K Means Example
Re-estimate Means



K Means Example
Re-assign Points to Clusters

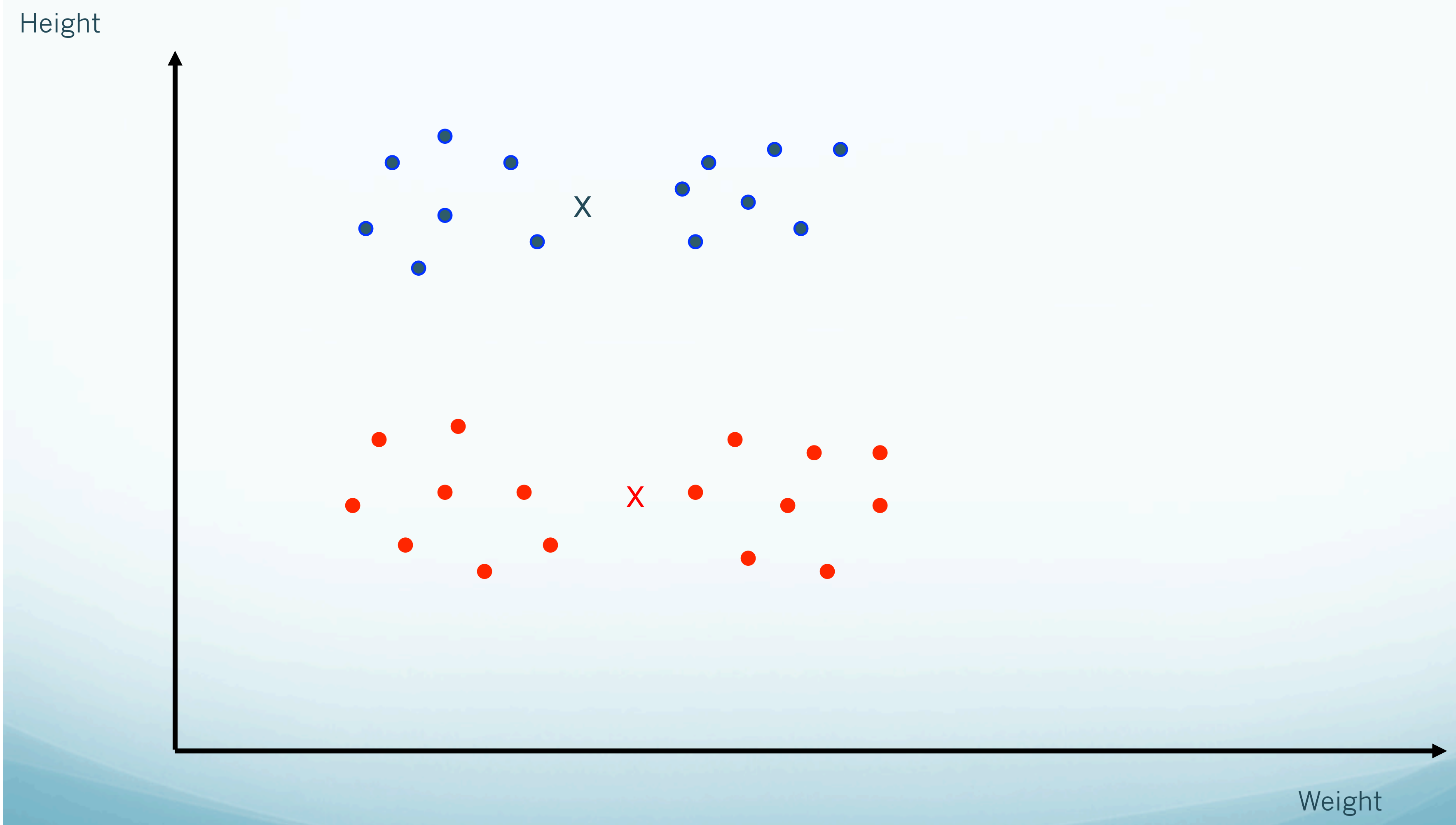


K Means Example
Re-estimate Means

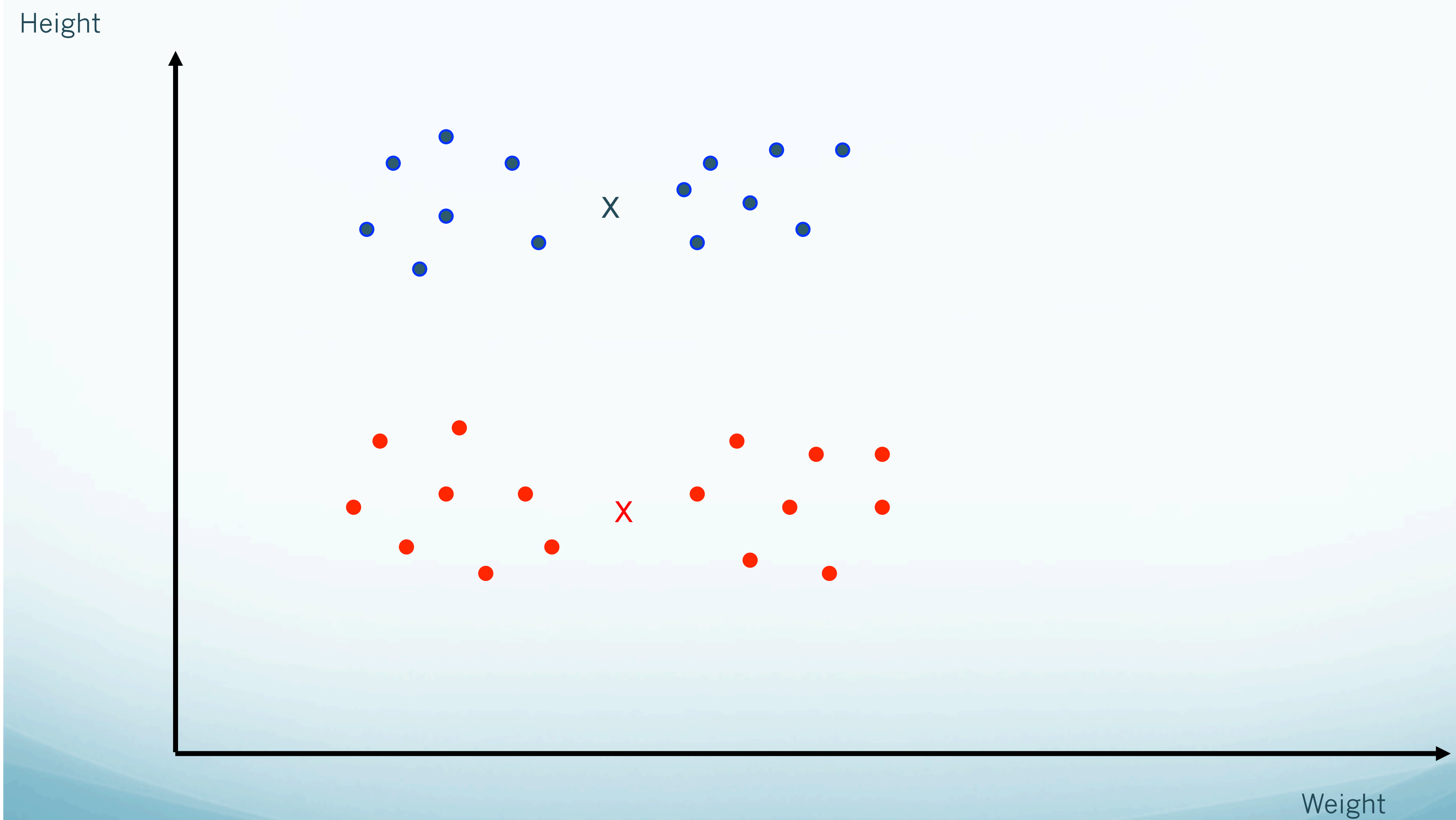


K Means Example

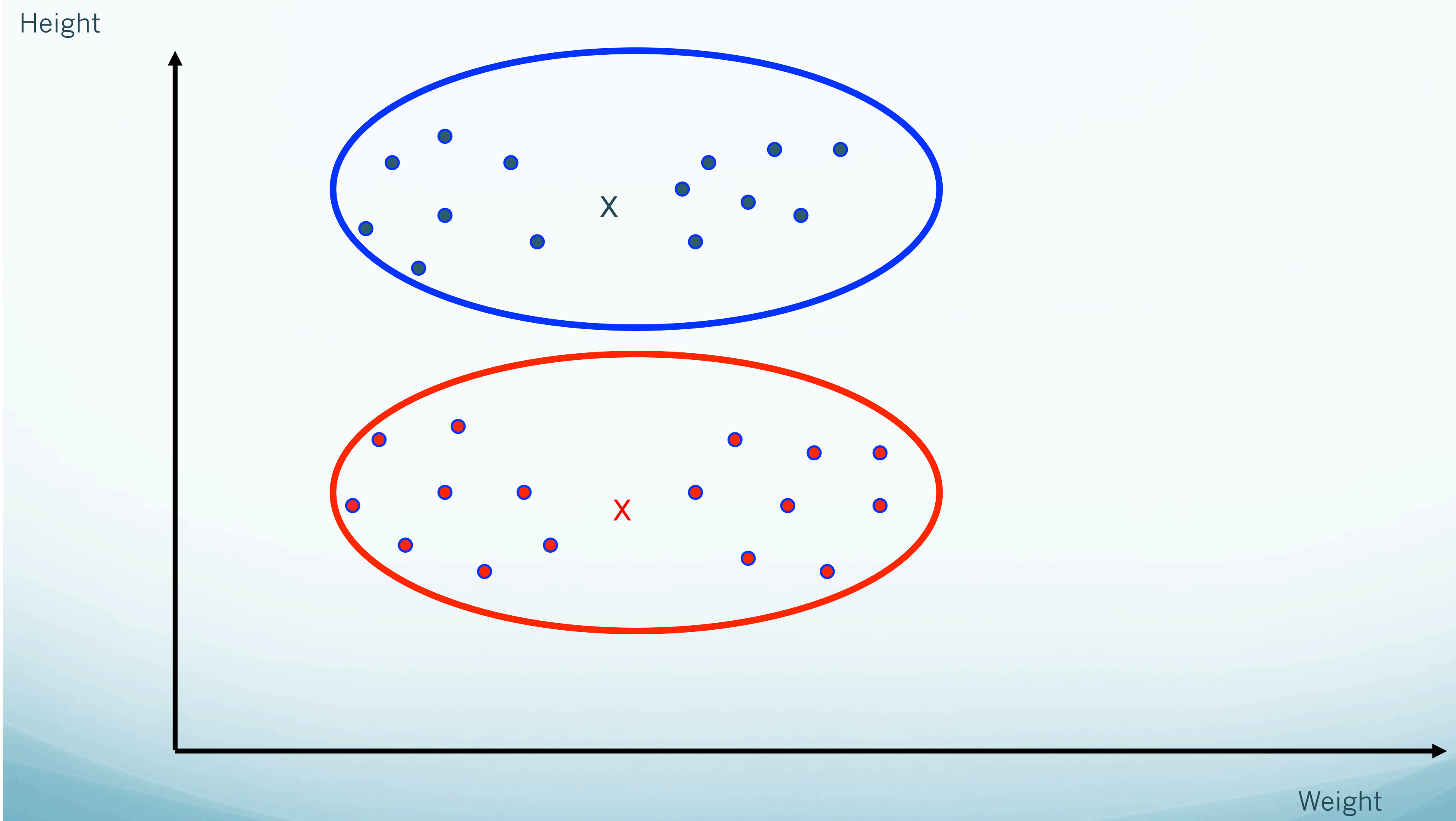
Re-assign Points to Clusters



K Means Example
Re-estimate Means and Converge



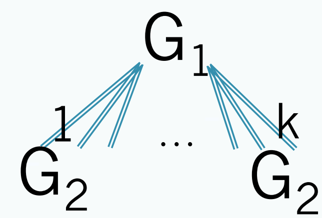
K Means Example
Convergence



Simple Observations

- Procedural algorithm effectively doing unconstrained tree search without BT
- Adding in constraints is trivial
  - Transitive closure for **together**
  - Not grow some parts of the tree for **apart** i.e. $G_1 = 1$
- AIJ Special Issue 2017 Article [Duong et al]

Search Tree

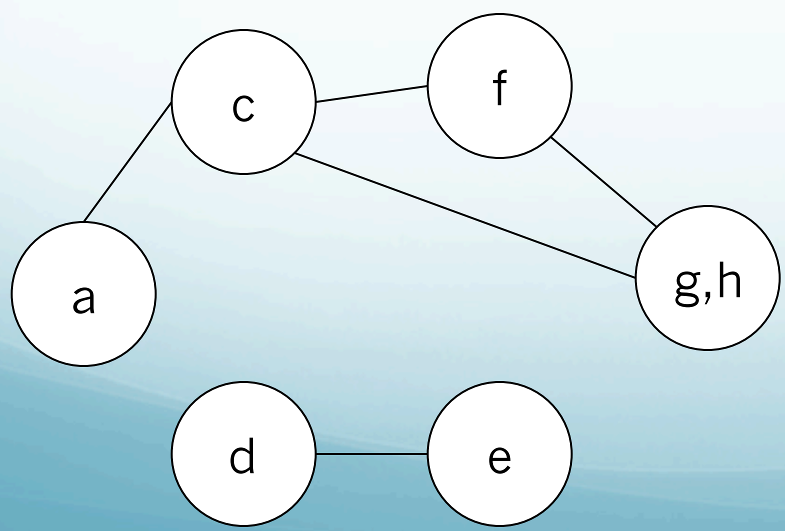


Feasibility Decision Sub-Problem

Given a set of data points *S*, a set of *together* and *apart* constraints, a lower (K_L) and upper bound (K_u) on the number of clusters, is there **at least one** single set partition of *S* into *k* blocks, $K_U \geq k \geq K_L$ such that no constraints are violated?

Result #1: Feasibility Problem for Just CL Constraints is NP-Complete

Instances a thru z
Constraints: ML(g,h) CL(a,c), CL(d,e), CL(f,g), CL(c,g), CL(c,f)



Some Conditions When Problem is Easy

Brooks' Theorem [Davidson and Ravi ACM KDD 2007]
Q-inductiveness of a Graph [Davidson and Ravi AAAI 2006]

Result #2: No Easy Work Arounds [Davidson Ravi ICML 2007]

- Quickly find a partition π that does **not** satisfy all constraints *C*
  - Work around #1: Repair π

Theorem 3.2 ***[No Efficient Repair Theorem]*** *The problem of repairing an infeasible clustering is* ***NP****-complete.*

- Work around #2: Minimally prune *C* to satisfy π
  - Constraint removal*:* minimum edge deletion k-coloring
  - *Constraint addition:* Start with the empty set and maximally retain constraints in *C* to obtain *C'* which π satisfies?

Theorem 3.3 ***[No Efficient Minimum Pruning]*** *For any $\rho \geq 1$, there is no ρ-approximation algorithm for the problem of deleting the minimum number of constraints, unless $\boldsymbol{P} = \boldsymbol{NP}$.*

Feasibility Results

Constraint	Given k	Unspecified k
Together	**P** [SDM05]	**P** [ECML05]
Apart	**NP-complete** [SDM05]	**P** [ECML05]
δ	**P** [SDM05]	**P** [ECML05]
ε	**P** [SDM05]	**P** [ECML05]
Together and ε	**NP-complete** [SDM05]	**P** [ECML05]
Together and δ	**P** [SDM05]	**P** [ECML05]
δ and ε	**P** [SDM05]	**P** [ECML05]
Together, Apart and ε	**NP-complete** [SDM05]	**NP-complete** [ECML05]

Davidson, Ian, and S. S. Ravi. "Using instance-level constraints in agglomerative hierarchical clustering: theoretical and empirical results." Data mining and knowledge discovery 18.2 (2009)

Davidson, Ian, and S. S. Ravi. "The complexity of non-hierarchical clustering with instance and cluster level constraints." Data mining and knowledge discovery 14.1 (2007): 25-61

2nd Generation Relaxed Versions of the Problem

Objective for spectral clustering
(Shi and Malik, 2000)

$$\underset{\mathbf{v}\in\mathbb{R}^N}{\operatorname{argmin}} \quad \mathbf{v}^T \bar{L} \mathbf{v},$$
$$\text{s.t.} \quad \mathbf{v}^T\mathbf{v} = 1, \; \mathbf{v} \perp D^{1/2}\mathbf{1}.$$

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} +1 & +1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 \\ -1 & -1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 \end{bmatrix}$$

Finding normalized min-cut

Lower-bounding how well the constraints in Q are satisfied

Normalizing v

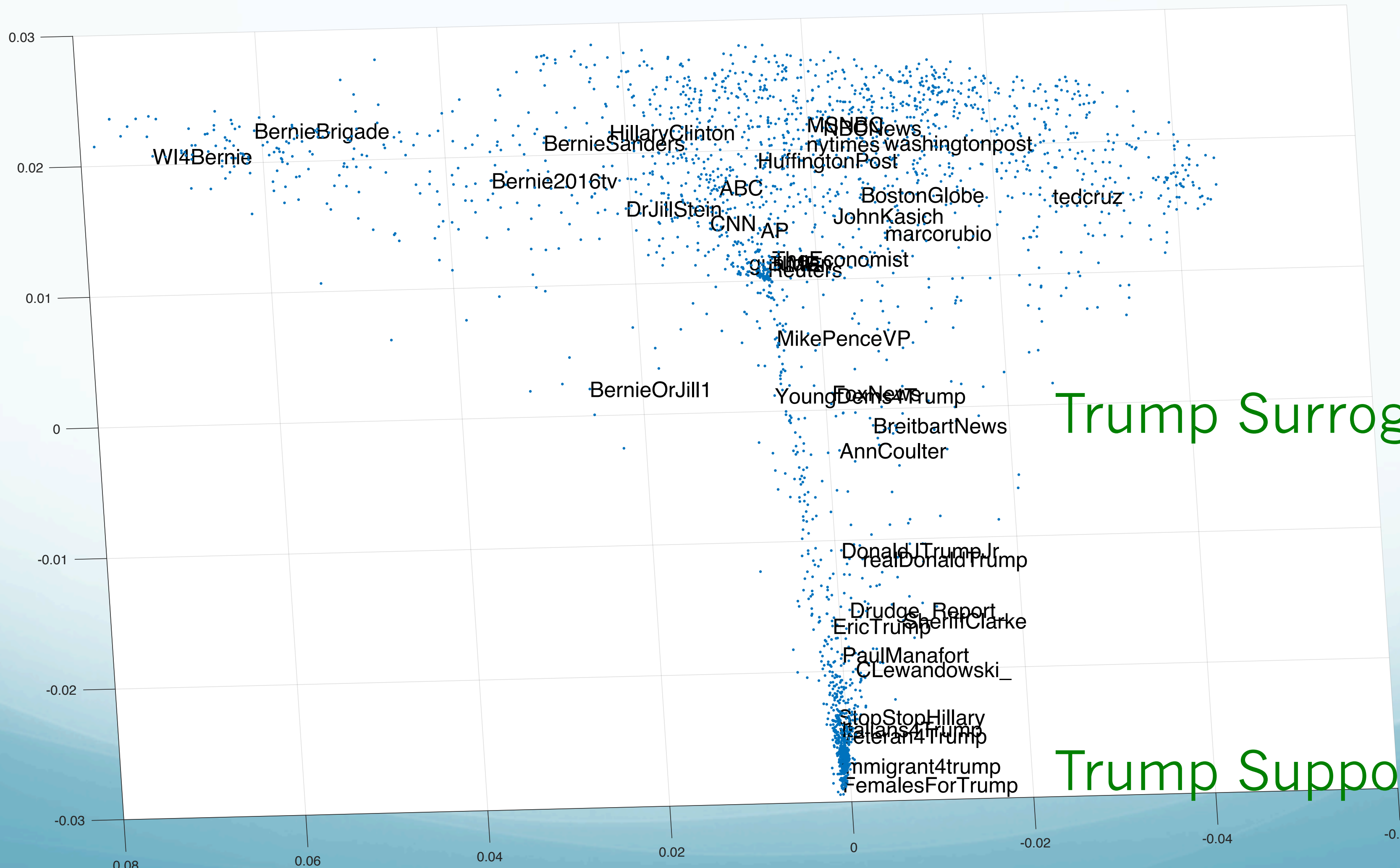
Ruling out the trivial solution

$$\boxed{\underset{\mathbf{v}\in\mathbb{R}^N}{\arg\min}\, \mathbf{v}^T \bar{L} \mathbf{v},} \; \text{s.t.} \; \boxed{\mathbf{v}^T \bar{Q} \mathbf{v} \geq \alpha,} \; \boxed{\mathbf{v}^T\mathbf{v} = \text{vol}(\mathcal{G}),} \; \boxed{\mathbf{v} \neq D^{1/2}\mathbf{1}.}$$

J. Shi and J. Malik. Normalized cuts and image segmentation. PAMI, 22(8):888–905, 2000.
X. Wang, B. Qian, I. Davidson. Flexible Constrained Clustering. ACM KDD 2010.
X. Wang, B. Qian, I. Davidson. On constrained spectral clustering and its applications. DMKD, 2014.

Cartography of Twitter- US Politics Retweets – Election

[Velcin et' al. SNAM 2019]



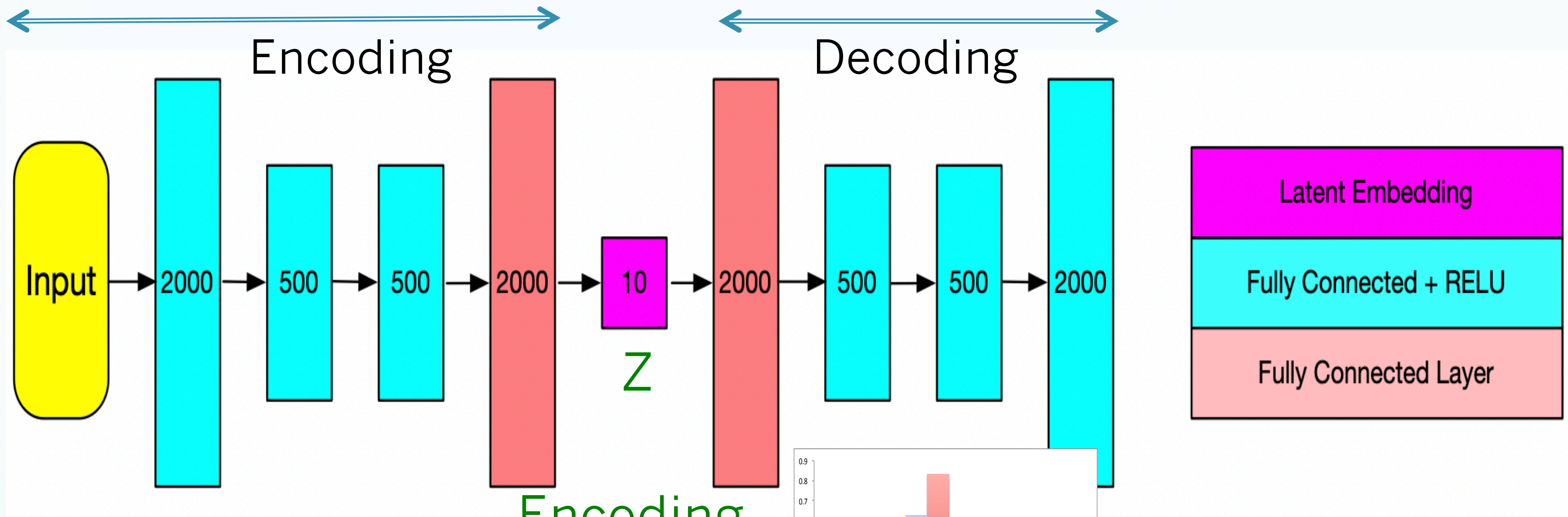
Trump Surrogates

Trump Supporters

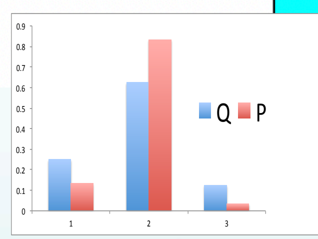
But Lots of Limitations

- 1: Limited to **conjunction** of **pairwise** constraints
  - Together/Apart, Easy to encode as a matrix
- 2: Relaxations means semantic meaning of constraints lost.
- 3: Scalability. All work by ourselves (KDD 10, DMKD 14) and others
  - Generalized eigen-value problem with time/space complexity O(n^3) where n is the number of entries in matrix
- 4: No inductive formulations

An Architecture To Perform Deep Embedding



Encoding Layer



Phase 1 – Auto encoder

Phase 2 – Refine encoding to represents an allocation indicator vector

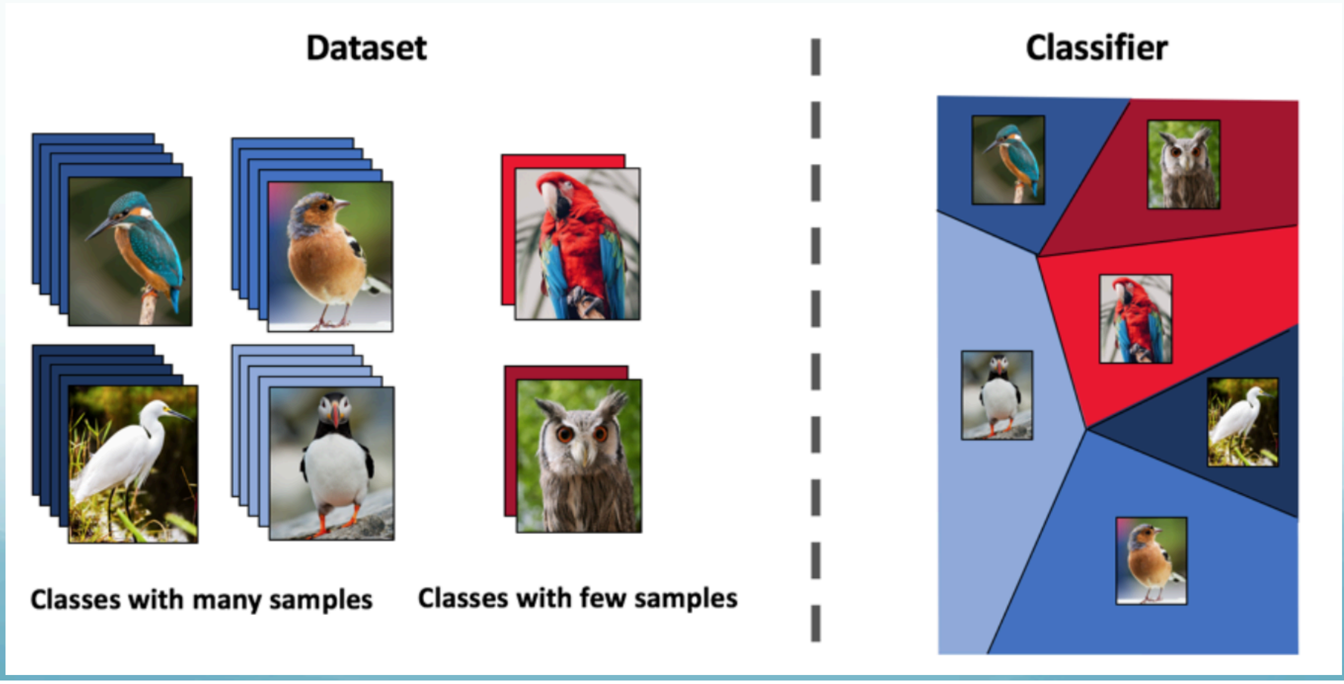
$$q_{ij} = \frac{(1 + ||z_i - \mu_j||^2/v)^{-\frac{v+1}{2}}}{\sum_{j'} (1 + ||z_i - \mu_{j'}||^2/v)^{-\frac{v+1}{2}}}$$

DOB Instance i is in class/cluster j

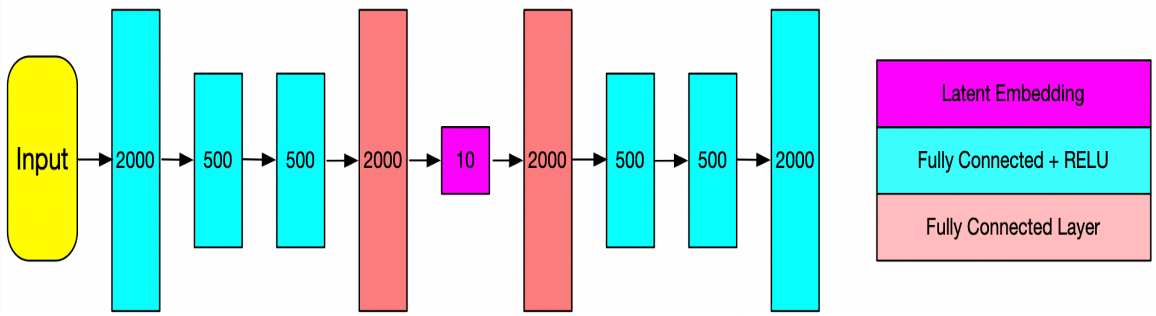
Xie, Junyuan, Ross Girshick, and Ali Farhadi. "Unsupervised deep embedding for clustering analysis." *International conference on machine learning*. 2016.

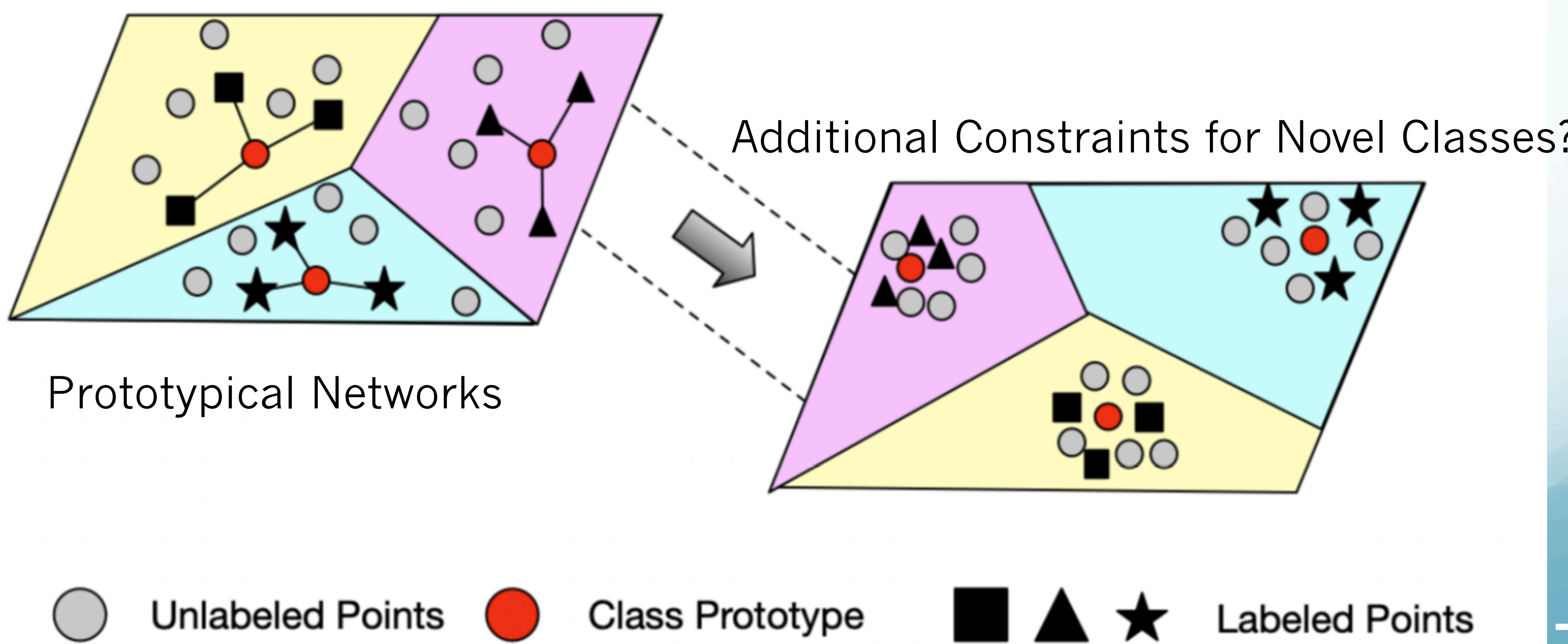
Application to Few Shot Learning

- Few shot learning: We do it our entire life
  - An 8 year old can recognize thousands of object categories … but we didn't have a training set for each!
  - Instead we transferred knowledge from known **base** classes/types to **novel** classes/types.



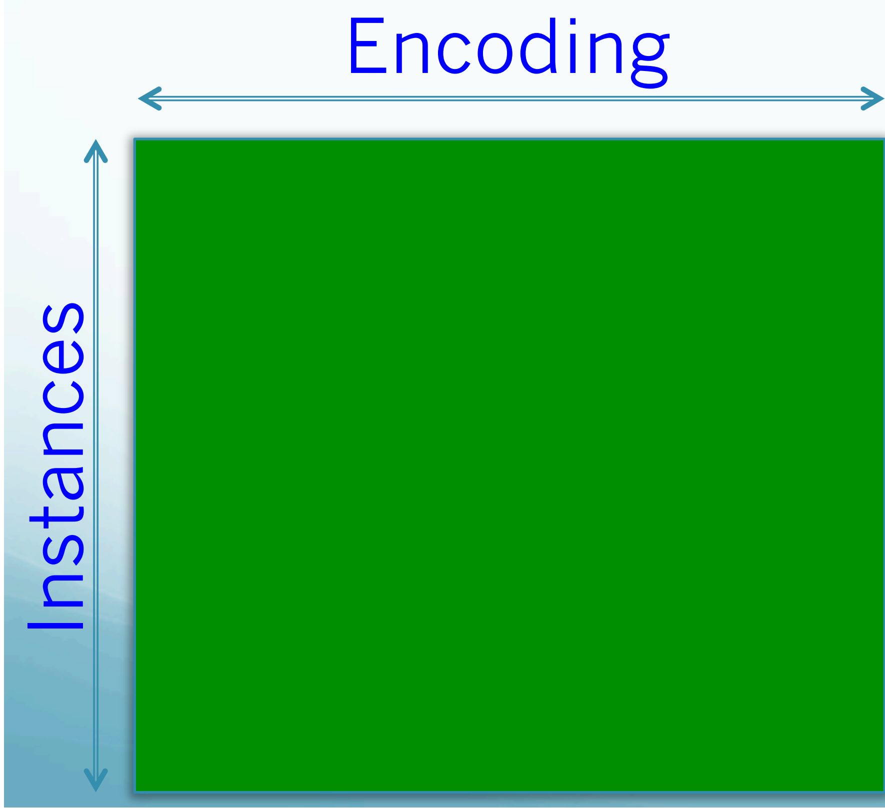
Prototypical Networks for Few Shot Learning

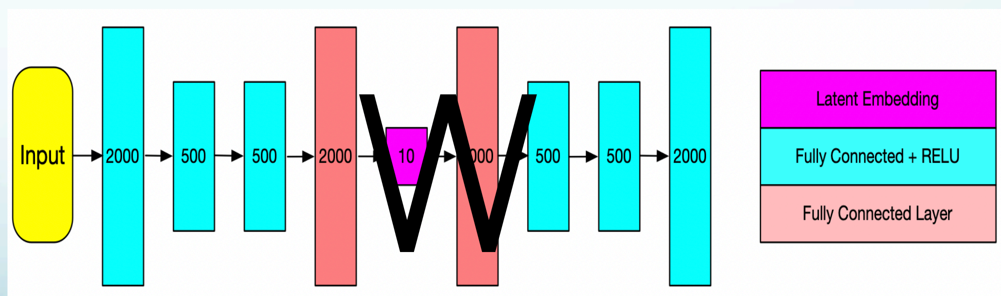




Idea of a Mini Batch Learning

$$\frac{\partial L_D}{\partial w} \approx \frac{\partial L_{D_1}}{\partial w} + \frac{\partial L_{D_2}}{\partial w} + \ldots \frac{\partial L_{D_m}}{\partial w}$$





Adding Constraints via New Differentiable Loss Functions

[Zhang, Basu, Davidson ECML 2019] and under preparation

Hard Instances

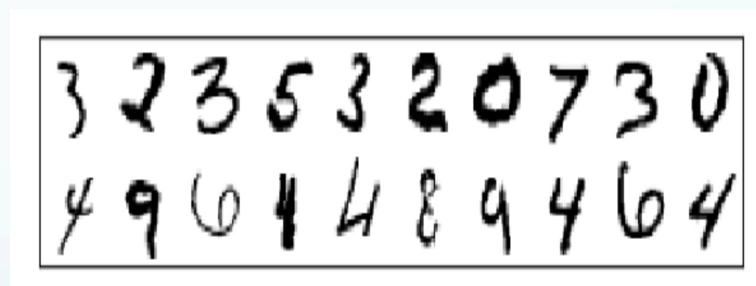
Easy Instances

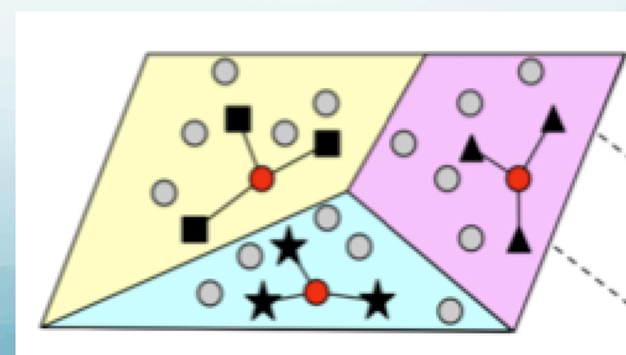
$$\ell_I = \sum_{t \in \{M_t < 0\}} M_t \sum_j {q_{tj}}^2 - \sum_{s \in \{M_s > 0\}} M_s \sum_j {q_{sj}}^2$$

Encoding

Instances

Instance Difficulty





Adding Constraints via New Differentiable Loss Functions

[Zhang, Basu, Davidson ECML 2019] and under preparation

Hard Instances

Easy Instances

$$\ell_I = \sum_{t\in\{M_t<0\}} M_t \sum_j {q_{tj}}^2 - \sum_{s\in\{M_s>0\}} M_s \sum_j {q_{sj}}^2$$

Encoding

Instance Difficulty

	MNIST	Fashion	Reuters
Acc(%)	88.29 ± 0.05	58.74 ± 0.08	75.20 ± 0.07
NMI(%)	86.12 ± 0.09	63.27 ± 0.11	54.16 ± 1.73
Epoch	87.60 ± 12.53	77.20 ± 11.28	12.90 ± 2.03

	MNIST	Fashion	Reuters
Acc(%)	91.02 ± 0.34	62.17 ± 0.06	78.01 ± 0.13
NMI(%)	88.08 ± 0.14	64.95 ± 0.04	56.02 ± 0.21
Epoch	29.70 ± 4.25	47.60 ± 6.98	9.50 ± 1.80

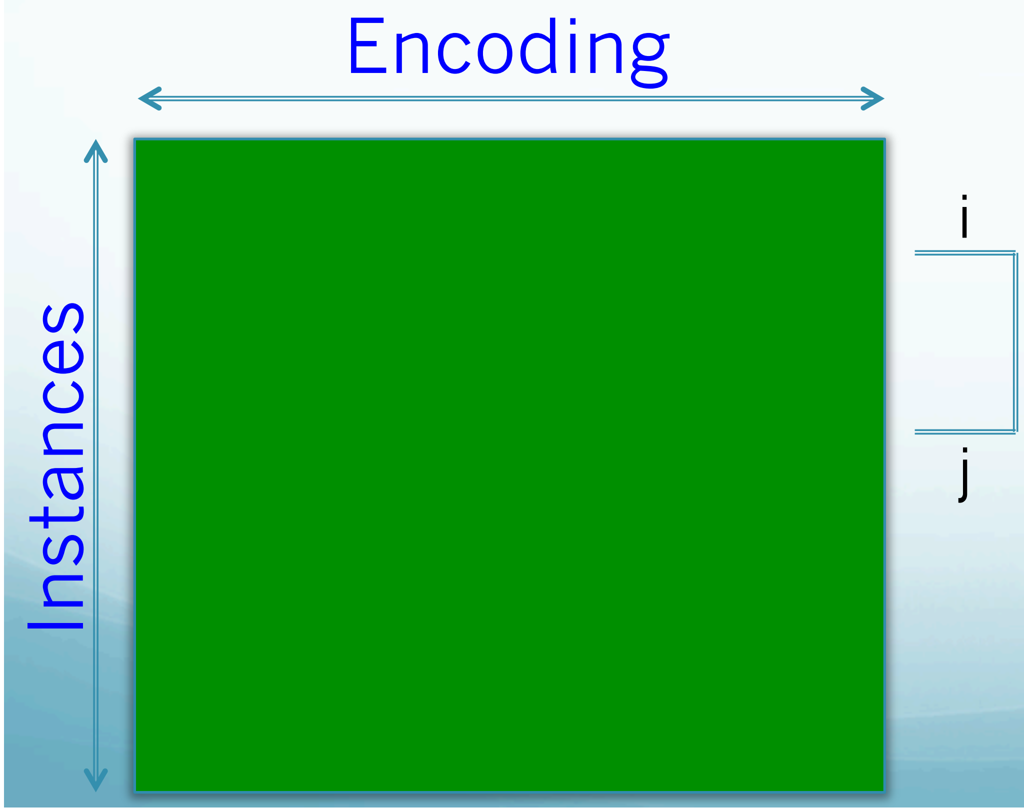
Table 1. Left table shows baseline results for Improved DEC [11] averaged over 20 trials. Right table lists experiments using instance difficulty constraints (mean $\pm$ std) averaged over 20 trials.

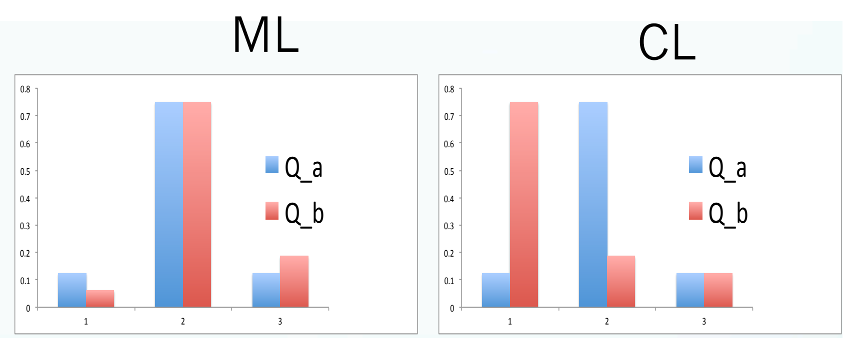
Weak Supervision via New Differentiable Loss Functions

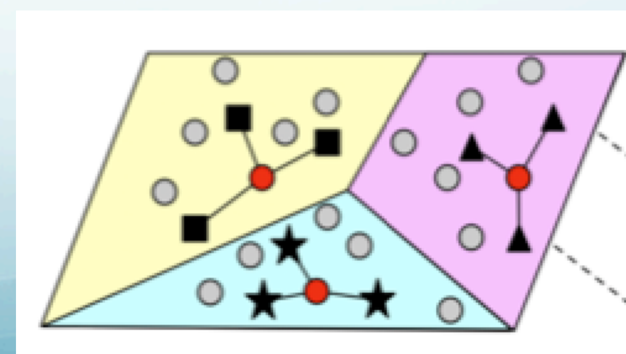
[Zhang, Basu, Davidson ECML 2019] and under preparation

$$\ell_{ML} = - \sum_{(a,b)\in ML} \log \sum_j q_{aj} * q_{bj}$$

$$\ell_{CL} = - \sum_{(a,b)\in CL} \log \left(1 - \sum_j q_{aj} * q_{bj}\right)$$







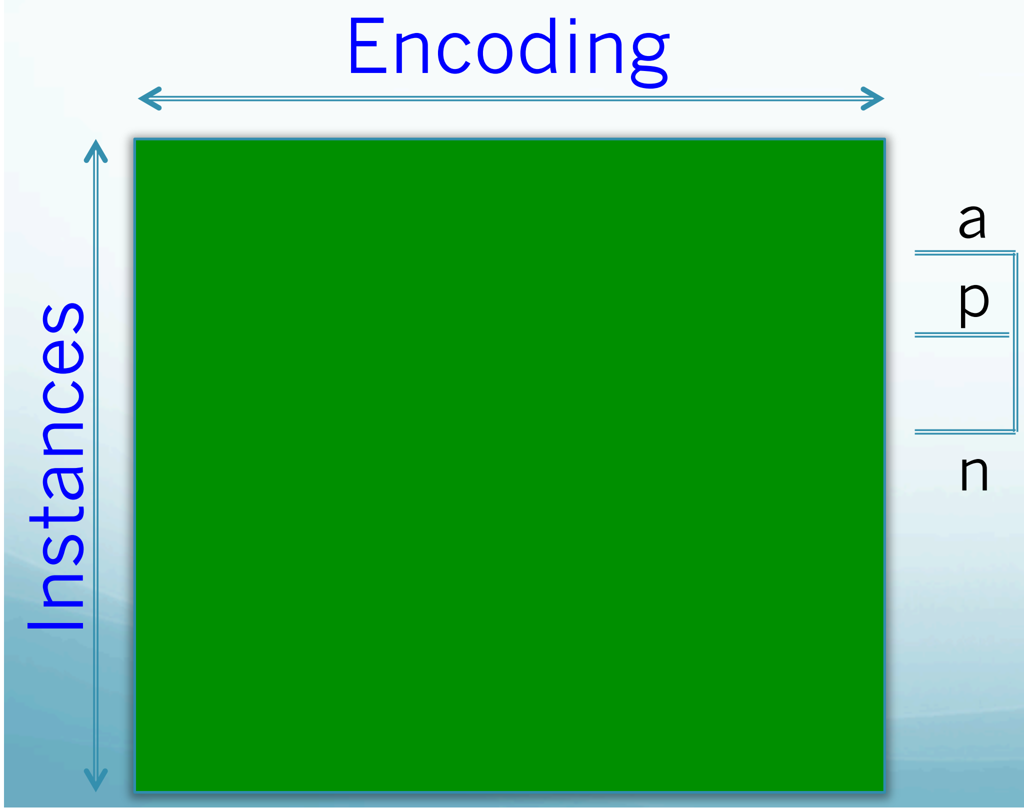
Adding Constraints via New Differentiable Loss Functions

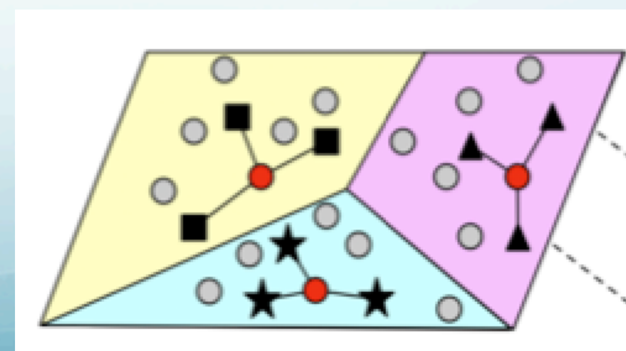
[Zhang, Basu, Davidson ECML 2019] and under preparation

$$d(q_a, q_b) = \sum_j q_{aj} * q_{bj}$$

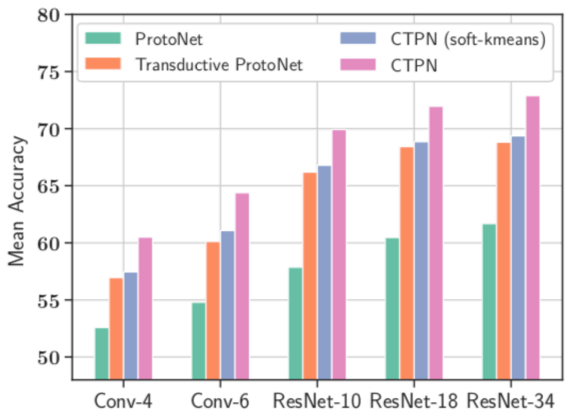
d is a match score

$$\ell_T = \sum_{(a,p,n)\in T} \max(d(q_a, q_n) - d(q_a, q_p) + \theta, 0)$$

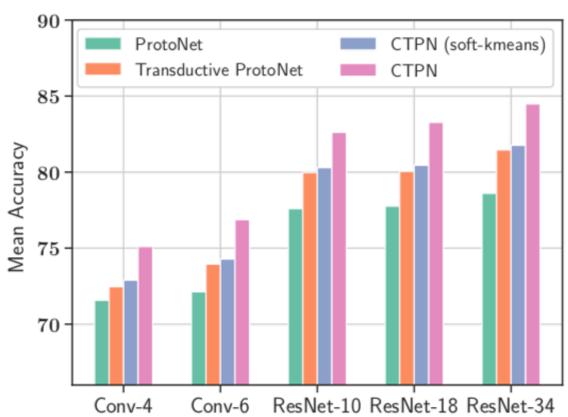




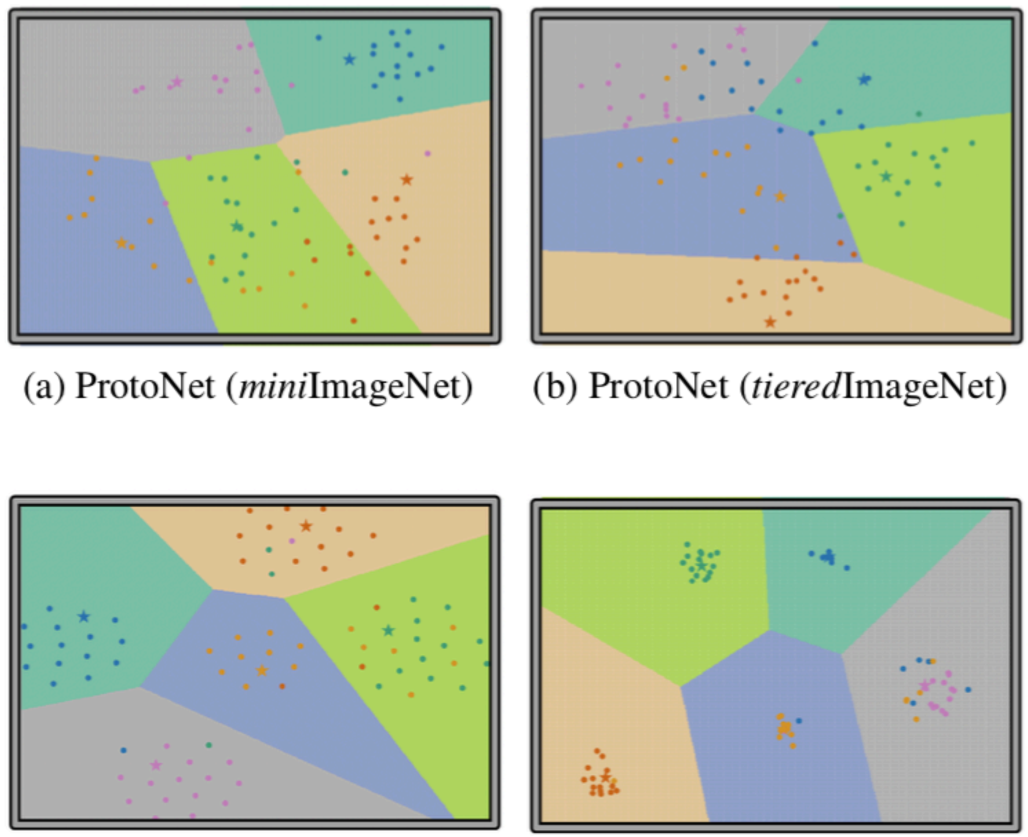
Some Results



(b) 1-shot on *tiered*Imagenet



(d) 5-shot on *tiered*Imagenet



(a) ProtoNet (*mini*ImageNet) (b) ProtoNet (*tiered*ImageNet)

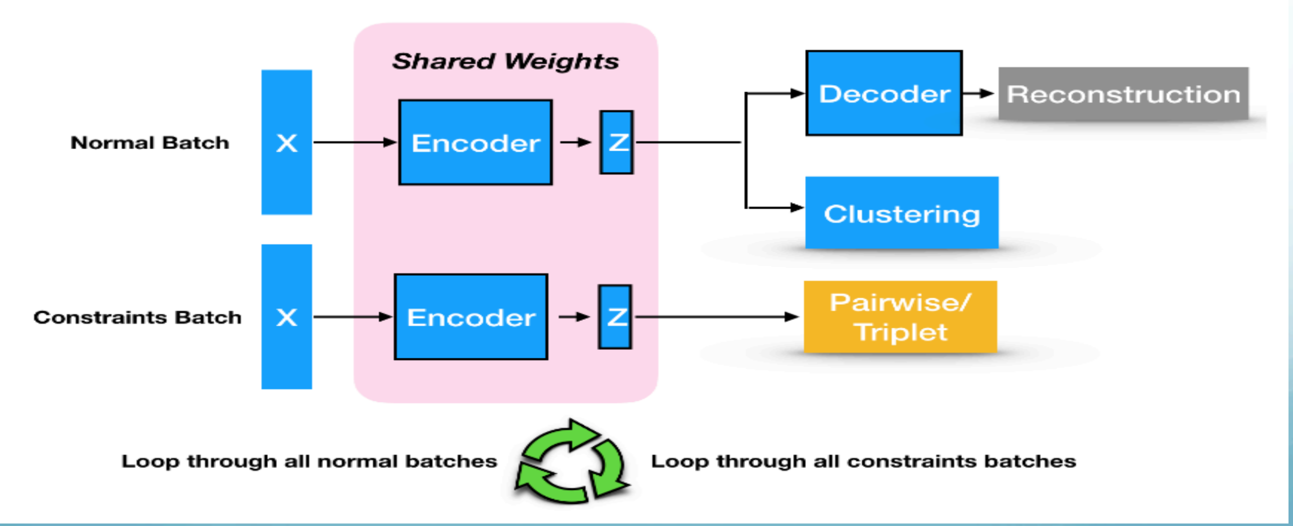
(c) CTPN (*mini*ImageNet) (d) CTPN (*tiered*ImageNet)

Figure 5: t-SNE latent embedding visualization.

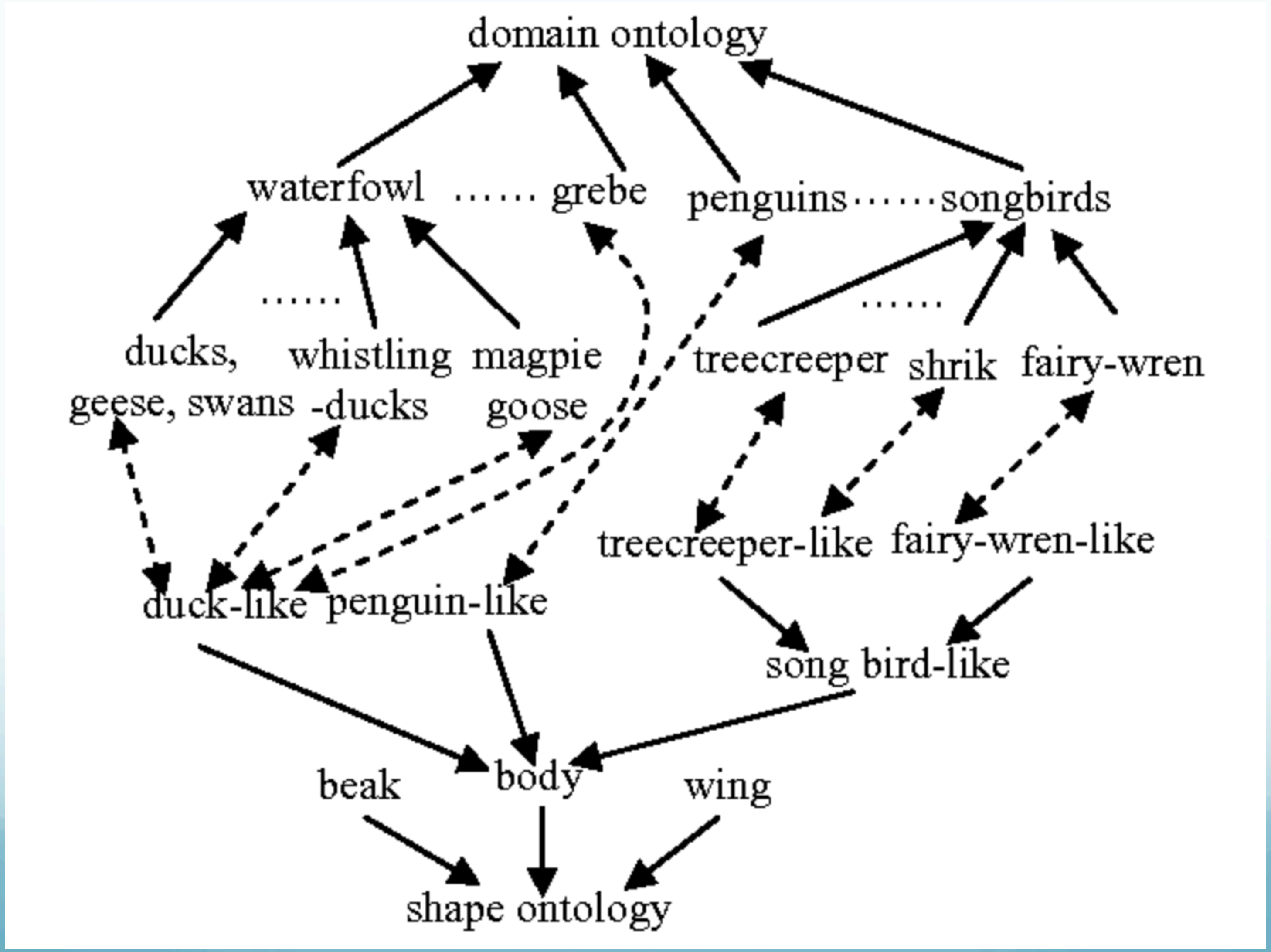
Take Home Message

[Code: https://github.com/blueocean92/deep_constrained_clustering]

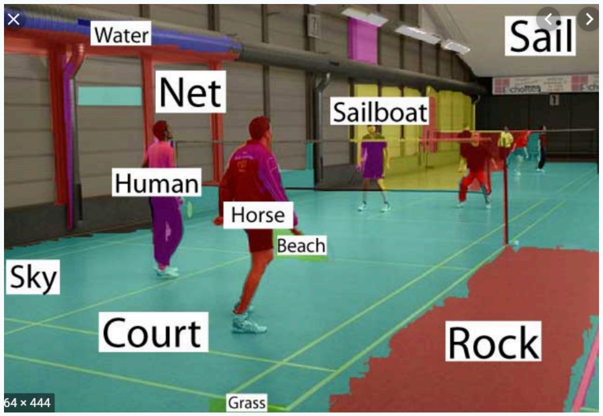
- Can add constraints **between** instance embeddings
- Add in new **differentiable** loss function
- Construction of mini-batch now becomes more challenging – compute transitive closure etc.
- Can't mix constraints, separate head for each type



Challenge 1- How To Encode An Ontology



Challenge 2 - Constraints Between Encoding/Output Not Instances



Encoding

$q_1, q_2, \dots q_m$

Instances

Each instance's encoding must satisfy

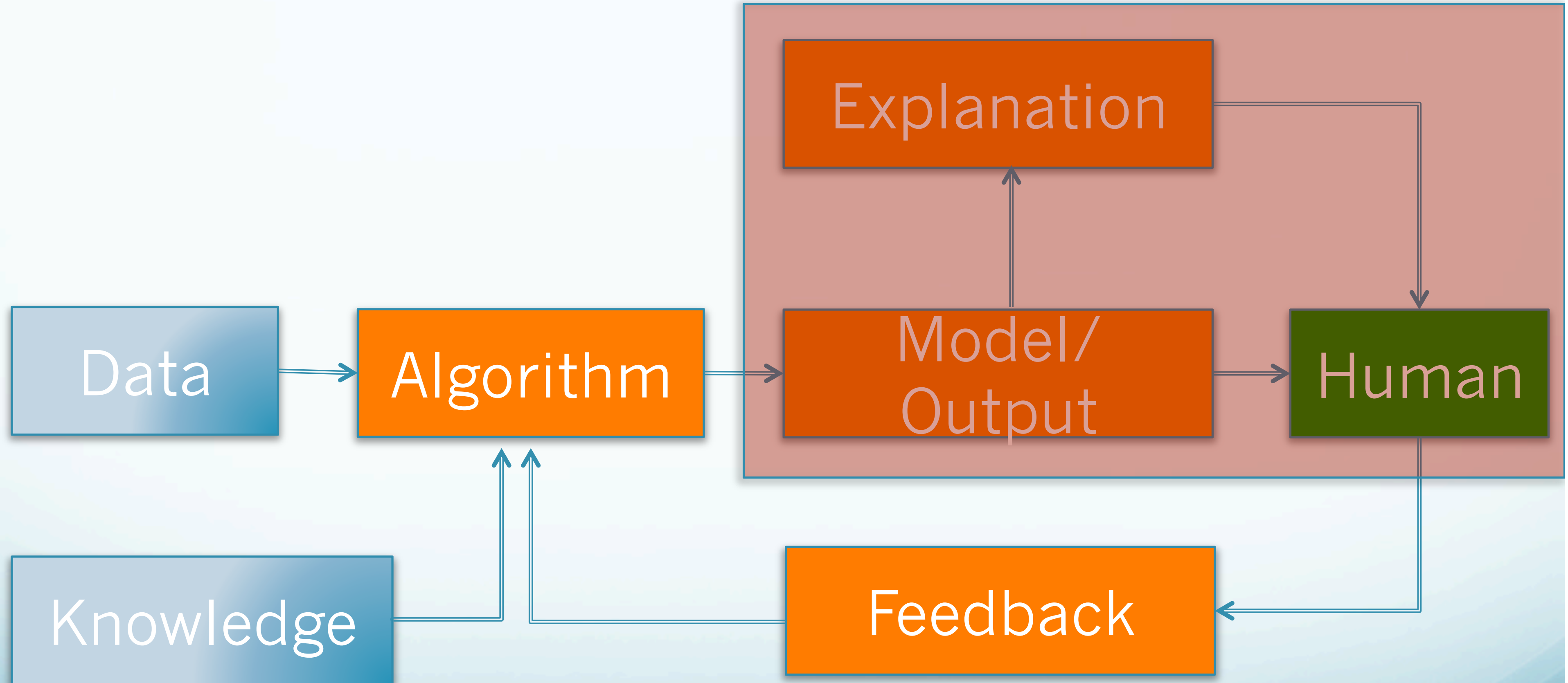
$$l \leq \sum_i q_i \leq u$$

$$q_{court} \wedge q_{net} \rightarrow \neg q_{rock} \wedge \neg q_{sail} \wedge \neg q_{sailboat}$$

Challenge 2 – Combining Symbolic and Sub-Symbolic AI

- Sub-symbolic: Deep network discovers “propositions” in complex data
- Symbolic: CP/SAT/ILP/etc. solver reasons about propositions
- Simple post-processing is a baby-step
- Need to directly use results of solver to adjust weights of DL
- Numerical challenge is computational paradigms are fundamentally different.

Adventure #2 – Extracting Explanations



Adventures in Formulating These Challenges as Constrained Optimization Problems

Explainable AI (XAI)

- Early work in ML/AI was fairly benign applications
  - Personalized hierarchies (Google), Personalized ranking of news (Yahoo!) etc.
- As the field matures, applications become controversial
  - Precision medicine for Schizophrenia treatment
- Demands the need for Explanation
  - Why: for validation, trust and fairness etc. to range of people: data scientist, domain experts, policy makers
- Two styles of XAI
  - New Algorithms: Simultaneously find a model that is also explainable
    - Pareto optimization formulations – IJCAI 2018
  - Existing Algorithms: Take the output of an existing algorithm and explain it – NIPS 2018

Learn Using One Set of Features Explain Using Another?

[The Cluster Explanation Problem: Complexity Results, Algorithms and Applications, Davidson et al. Neurips 18]

- Features used for learning are not interpretable
  - Deep learning representations (Word2Vec, Bert, AE)
  - Graphs i.e. social networks
- Features for learning are private/sensitive
- Features used to obtain learning are no longer available (historical clusters)
  - Electoral district maps

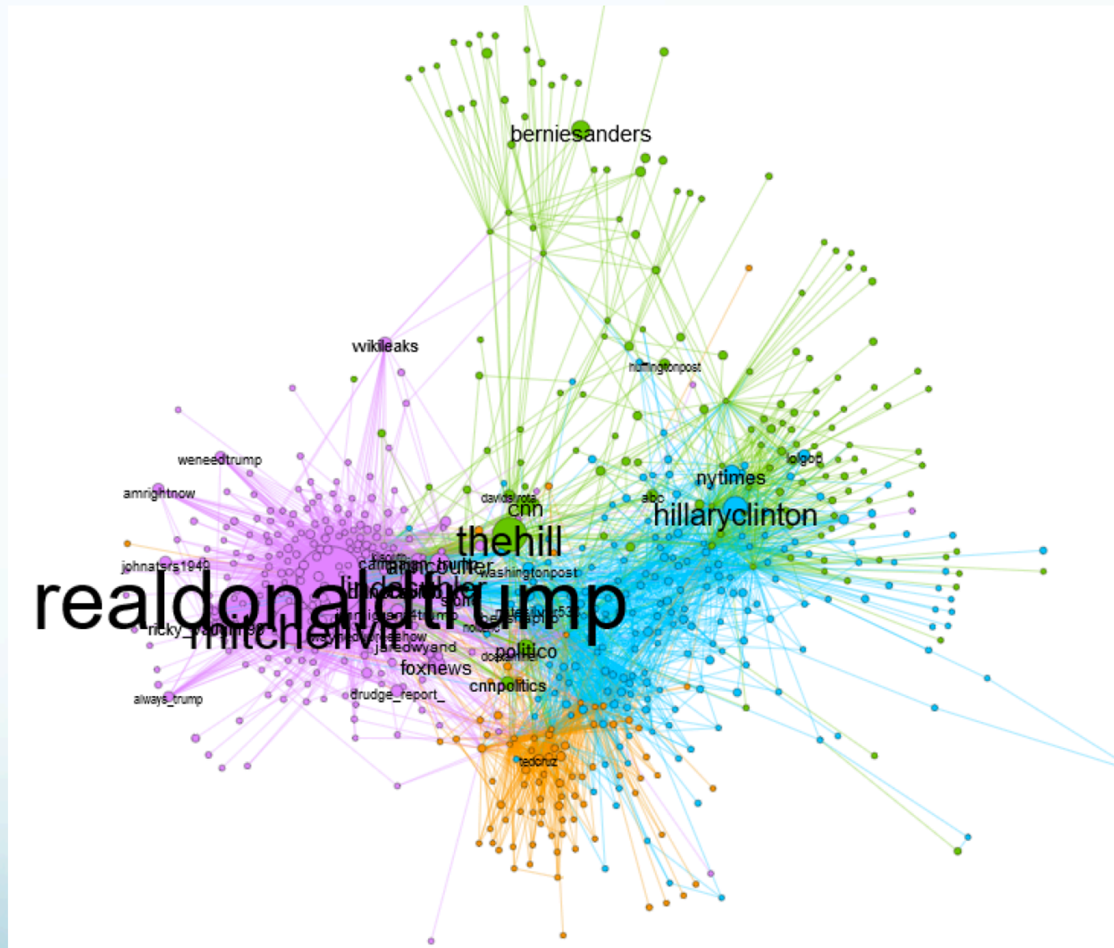
Twitter Data from ERIC Lab Univeristy Lyon - 2

- Election Tweets (**French and USA**) from 01/01/2016 to 22/08/2016
  - Covers the primary season for the USA
- Communities formed based on **follower network**
- Explain communities using **hashtag usage**

HASHTAG

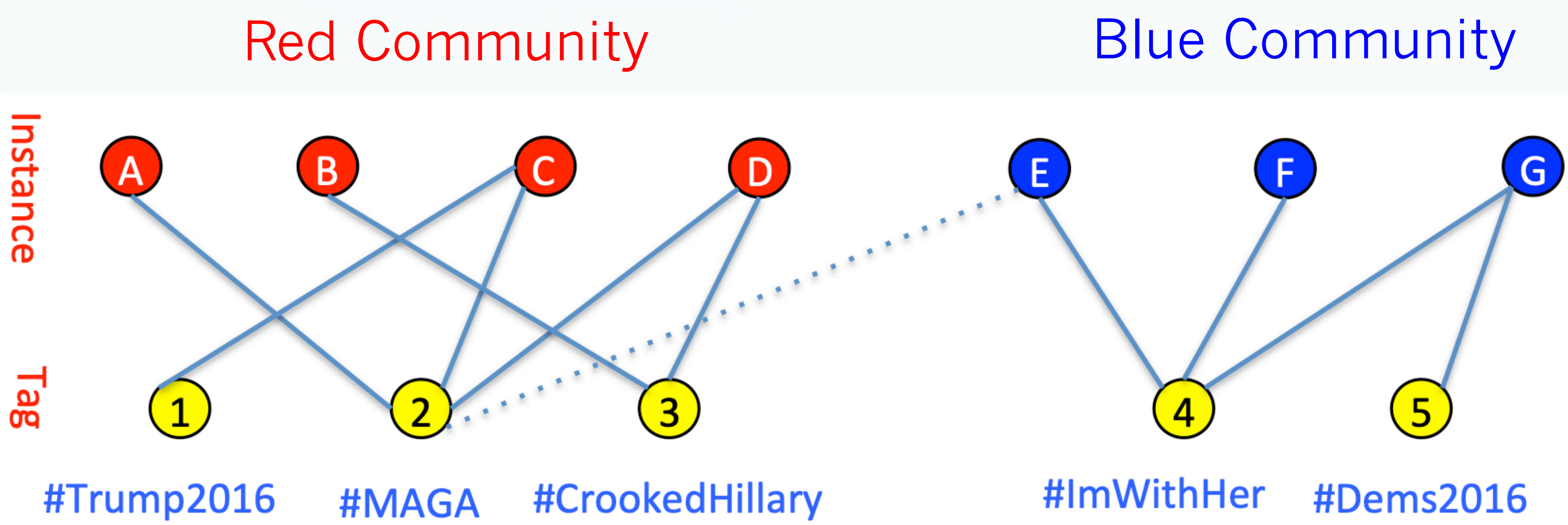
Segmenting Graphs

- Many classic algorithms – i.e. Louvain method
- Finds useful communities
- But doesn't explain what they are talking about.



Modeling the Explanation Problem as a Bipartite Graph

A Simple Example with Just Two Clusters
(ignore dotted line)



Explanation Problem (informally): Pick the smallest subset of yellow nodes/tags that "cover" all red instances. Chose a **different** subset of yellow nodes/tags for the blue tags

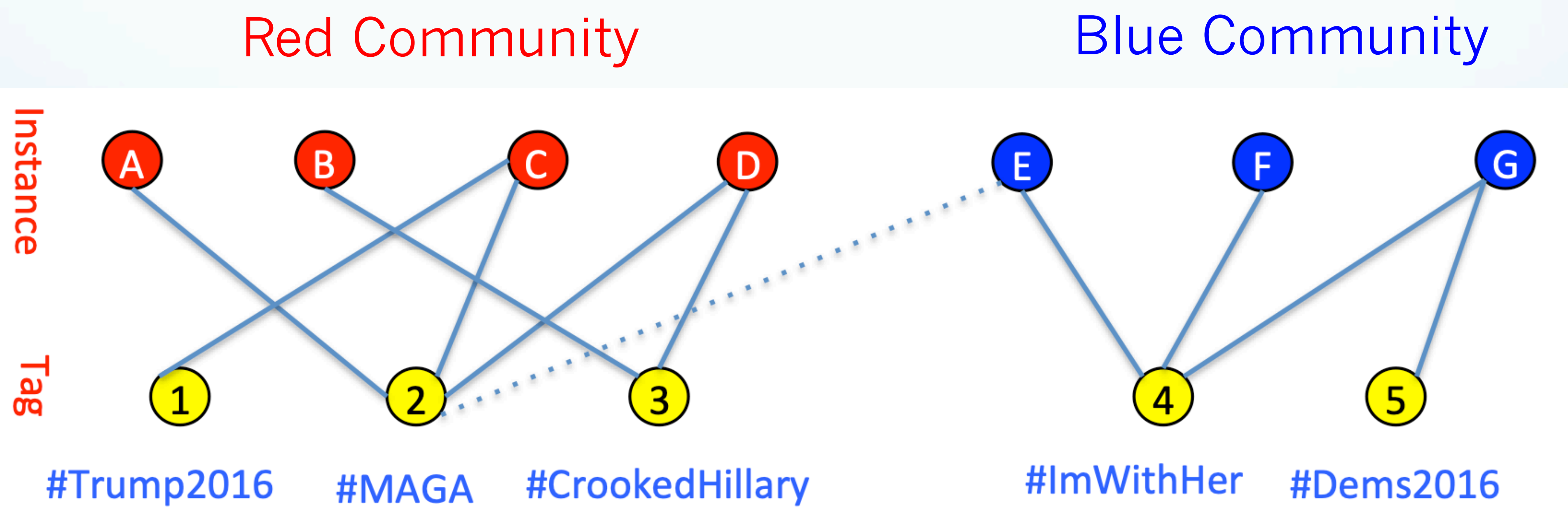
Formalizing DTDF Problem

[Disjoint Tag Descriptor Feasibility Problem]

The goal is to find a subset $T_j \subseteq T$ of tags for each cluster C_j $(1 \leq j \leq k)$ such that all the following conditions are satisfied.

(a) For each cluster C_j and each item $s_i \in C_j$, T_j has at least one of the tags in t_i; formally, $|T_j \cap t_i| \geq 1$, for each $s_i \in C_j$ and $1 \leq j \leq k$.

(b) The sets $T_1, T_2, \ldots, T_k$ are pairwise disjoint.



Formalizing This As An ILP

A triple set to 1 iff instance *i* is in cluster *k* and uses tag *j*

Find a simple explanation

Cover each and every instance

Upper bound overlap

$$argmin_X \sum_{i,j} X_{i,j}$$

$$s.t. \quad \sum_j X_{k,j} S^k_{i,j} \;\geq\; 1 \;\; \forall\, i \in C_k, \;\; \forall\, k$$

$$s.t. \quad \sum_i X_{i,j} \;\leq\; w_j \;\; \forall\, j$$

This is very similar to the set cover problem, one of Karp's 21 original intractable problems.

Formalizing This As An ILP

$$\begin{aligned} & argmin_X \sum_{i,j} X_{i,j} \\ s.t. \quad & \sum_j X_{k,j} S^k_{i,j} \ \geq \ 1 \ \ \forall \ i \in C_k, \ \forall \ k \\ s.t. \quad & \sum_i X_{i,j} \ \leq \ w_j \ \ \forall \ j \end{aligned}$$

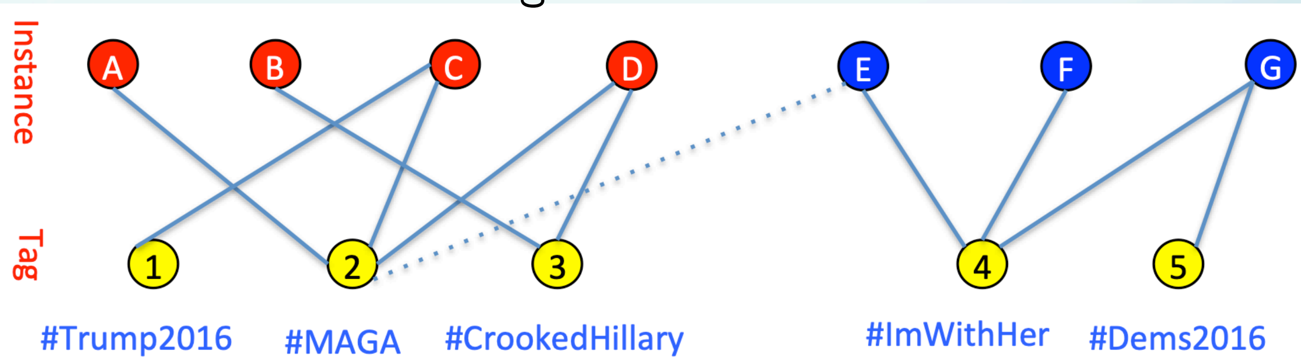
Find a simple explanation

Cover each and every instance

Upper bound overlap

Solution?

Ignore the dotted line



Formalizing This As An ILP

$$argmin_X \sum_{i,j} X_{i,j}$$

Find a simple explanation

$$s.t. \quad \sum_j X_{k,j} S^k_{i,j} \; \geq \; 1 \;\; \forall \, i \in C_k, \; \forall \, k$$

Cover each and every instance

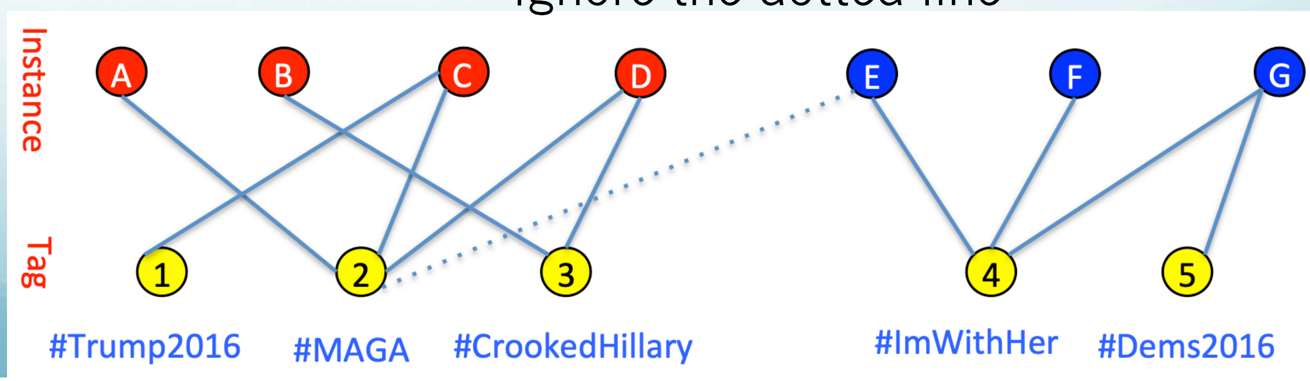
$$s.t. \quad \sum_i X_{i,j} \; \leq \; w_j \;\; \forall \, j$$

Upper bound overlap

Optimal Solution
X_R = {MAGA, CrookedHillary}
X_B = {ImWithHer}

But if E only has the MAGA tag, and
w_j = 1 \forall j

Ignore the dotted line



Formalizing This As An ILP

$$argmin_X \sum_{i,j} X_{i,j}$$

Find a simple explanation

$$s.t. \sum_j X_{k,j} S^k_{i,j} \geq 1 \;\; \forall\, i \in C_k, \;\forall\, k$$

Cover each and every instance

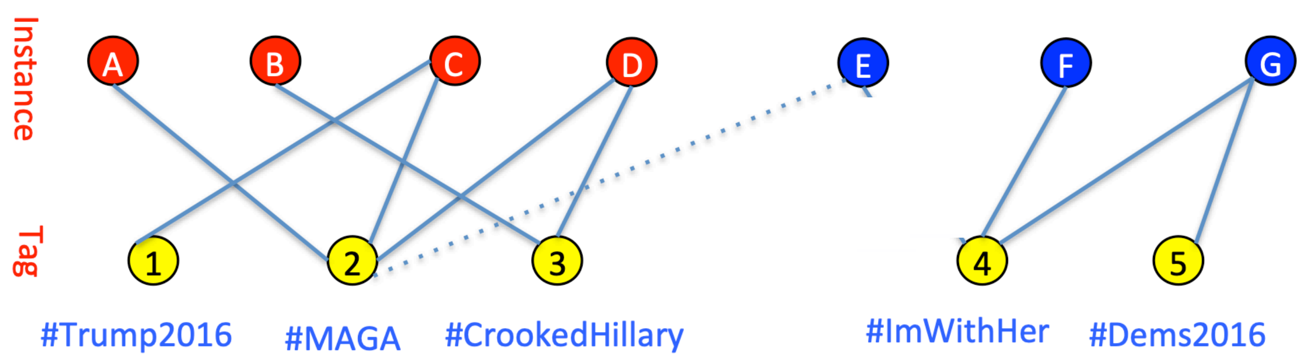
$$s.t. \sum_i X_{i,j} \leq w_j \;\; \forall\, j$$

Upper bound overlap

Optimal Solution?

But if E only has the MAGA tag,
w_j = 1 \forall j

no feasible solution exists



Variation #1

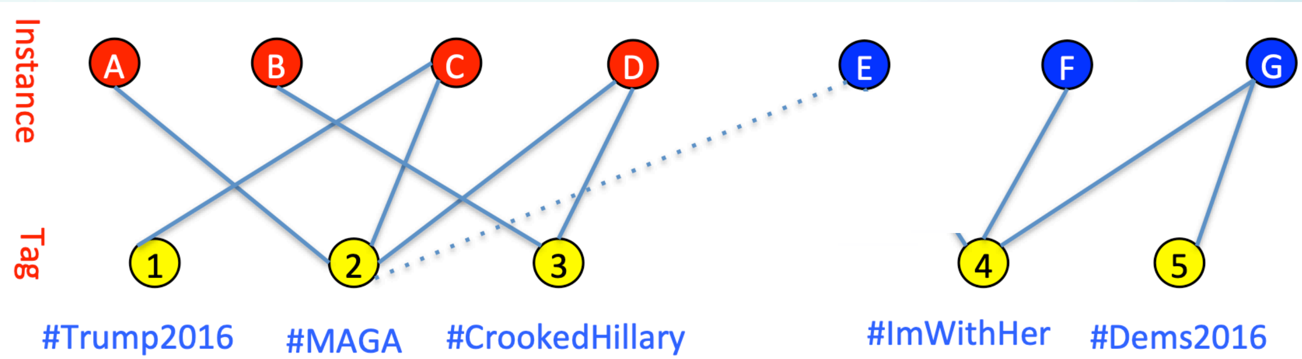
- Cover or forget (constraint replacement)

$$s.t.\ \ z_i + \sum_j X_{k,j} S^k_{i,j} \geq 1 \ \forall\, i \in C_k,\ \forall\, k$$

$$s.t.\ \sum_i z_i \leq I_k \ \forall i \in C_k,\ \forall\, k$$

Optimal Solution
X_R = {MAGA, CrookedHillary}
X_B = {ImWithHer}
Z_5 = 1

If E only has the MAGA tag



Variation #2

- Composition constraints

$$s.t.\ \ X_{k,i} + X_{k,j} \leq 1\ \ \forall\ \{i,j\} \in \texttt{Apart},\ \forall\ k$$ (#MAGA, #Clinton)

$$s.t.\ \ X_{k,i} = 1 \rightarrow X_{k,j} = 1\ \ \forall\{i,j\} \in \texttt{Together},\ \forall\ k$$ (#MAGA, #Trump)

Lemma 1

Any disjunction of literals $v_1 \ldots v_m$ can be represented by a linear inequality (i.e. $v_1 \vee v_2 \ldots \vee v_m \equiv \sum_i \mathbf{v}_i \geq -m + 2$).

Lemma 2

Any conjunction of literals $v_1 \ldots v_m$ can be represented by a linear equality (i.e. $v_1 \wedge v_2 \ldots \wedge v_m \equiv \sum_i \mathbf{v_i} = m$).

Theorem 3

Given a set of literals $v_1 \ldots v_m$, any set of clauses using those literals in conjunctive normal form can be represented by a system of linear inequalities: $A_{=}\mathbf{v} = \mathbf{b}_{=}, A_{\geq}\mathbf{v} \geq \mathbf{b}_{\geq}$.

TrumpTrain $\vee$ MAGA =>
Not(ImWithHer $\vee$ Clinton)

AAAI 13 Paper of Ours

But the Explanation Problem is Intractable

Theorem 3.1 *The* DTDF *problem is* ***NP****-complete even when the number of clusters is 2 and the tag set of each item is of size at most 3.*

What were the three options if the problem is intractable?.

Contributions

Theorem 3.1 *The* DTDF *problem is* ***NP****-complete even when the number of clusters is 2 and the tag set of each item is of size at most 3.*

What were the three options if the problem is intractable?

If we require:

i) each explanation must have at most α tags

ii) no two explanations may have more than β tags in common.

Theorem 5.1 *The* (α, β)-CONS-DESC *problem can be solved in polynomial time when the number of clusters k is fixed. This algorithm can also handle* `Together` *and* `Apart` *composition constraints.*

This is called a fixed parameter tractable solution

Simple Fixed Parameter Tractable Algorithm

Algorithm 1: Description of our Algorithm for (α, β)-CONS-DESC

Input : A collection of k clusters $C_1, C_2, \ldots, C_k$ with tag sets for each instance in each cluster.
Output : A valid descriptor with at most α tags for each cluster such that any pair of descriptors have at most β tags in common. (Please see the main text for the definition of a valid descriptor.)

for Cluster C_1 **do**
  Get the next valid descriptor D_1.
  for Cluster C_2 **do**
    Get the next valid descriptor D_2.
      ⋮
    for Cluster C_k **do**
      Get the next valid descriptor D_k.
      Let $\mathcal{D} = (D_1, D_2, \ldots, D_k)$.
      if *Each pair of descriptors in $\mathcal{D}$ have at most β tags in common* **then**
        Output $\mathcal{D}$ as the solution and **stop**.
      end
    end
  end
end
Print "No solution".

Let N be the most tags used in a cluster

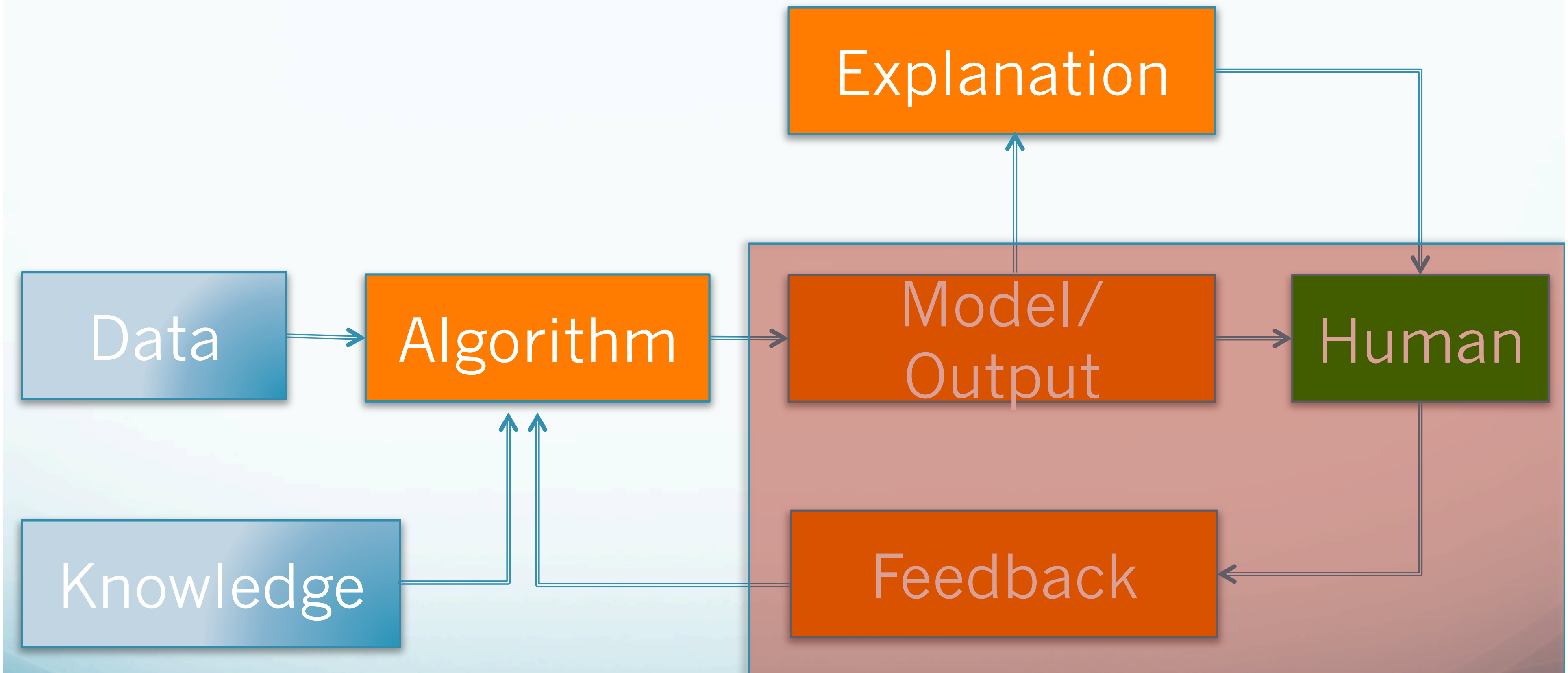
Possible Explanations N Choose α = $O(N^{\alpha k})$

$|T\_i| \leq \alpha$ hence $O(\alpha^2)$ to check constraints
Trivial to check for β overlap

Future Work

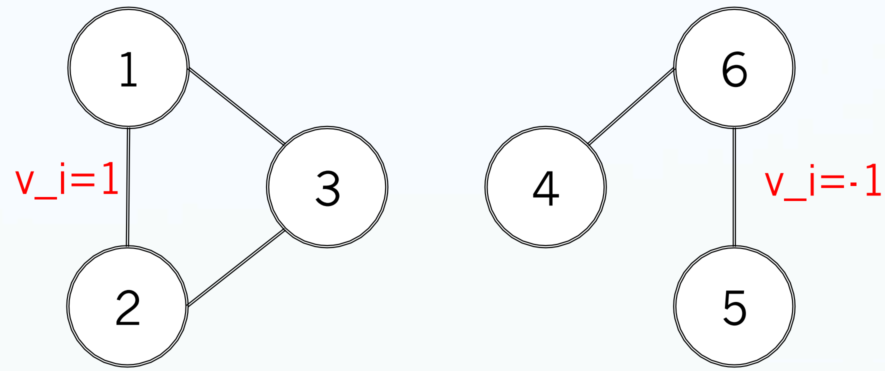
- https://web.cs.ucdavis.edu/~davidson/description-clustering/
  - See readme file
- **Cleverer algorithms:** The algorithm is polynomial, but brute force search
  - Can we use a clever branch and bound method
- **More Descriptive Explanations**
  - Meta information about the tags
  - Lots of other types of explanations beyond disjunctions
    - CNF, DNF
- **Explanations Using other Information**
  - Explanations using other types beyond tags
  - Need to use SMT, OMT, CP solvers not ILP solvers
- Measuring **stability of explanations** for measures of trust

Adventure #3 – Using Constraints as a Dialog Mechanism



Adventures in Formulating These Challenges as Constrained Optimization Problems

Consider our Dinner Party Problem Again

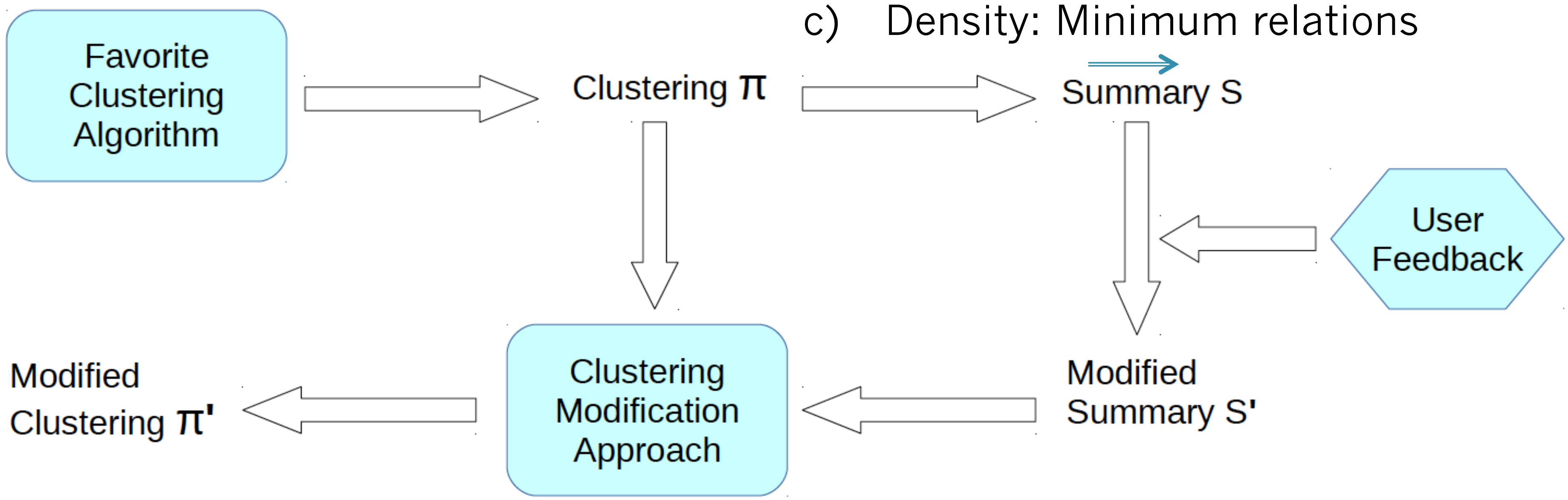


- After the division maybe we find the result useful but perhaps group i has:
  - Mostly males
  - Wide range of ages
  - No common hobbies etc.
- Wish to modify group i to make it better but not unduly change existing groups.

Minimal Modification Problem

Framework for Minimal Clustering Modification via Constraint Programming, Tom Kuo et. al. AAAI 17

a) Cardinality: Balance males/females
b) Geometric: Diameters
c) Density: Minimum relations



Minimally modify Π to obtain Π' to satisfy S'

$$\begin{array}{ll} \underset{\Pi'}{\textit{minimize}} & d(\Pi, \Pi') \\ \textit{subject to} & \Pi' \textit{ satisfies } S' \end{array}$$

Intractable Problem

A Framework for Minimal Clustering Modification via Constraint Programming, Tom Kuo et. al. AAAI 17

Theorem (1)

The reclustering problem where $\ell = 2$ is NP-complete.

Proof idea: reduction to Covering Points by Unit Squares.

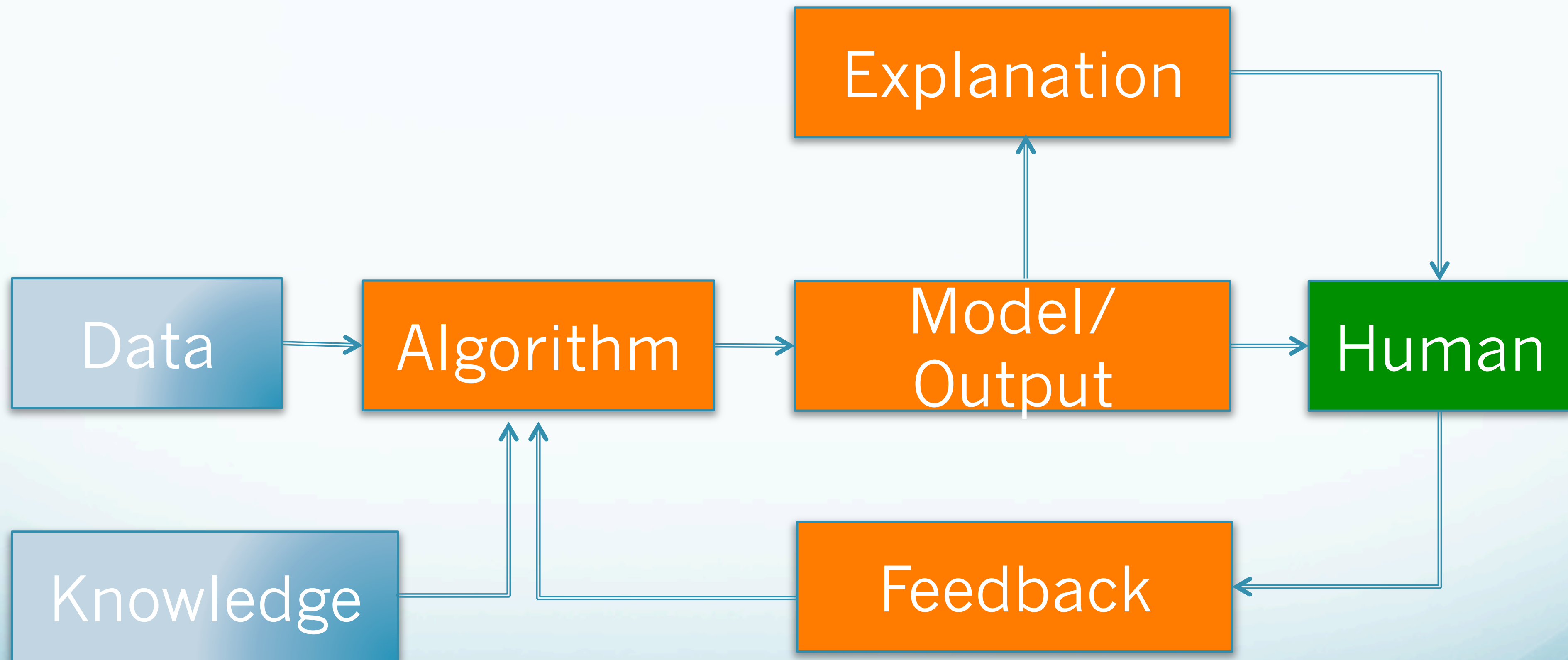
Even for very limited settings

Theorem (2)

Suppose the number of dimensions along which the maximum diameter must be reduced is a variable ℓ. The reclustering problem is NP-complete for any $k \geq 3$.

Proof idea: similarly reduction to Covering Points by Unit Hypercubes.

The Three Adventures



Adventures in Formulating These Challenges as Constrained Optimization Problems

Thank You, Questions and Acknowledgements Again

- Lots of Smarter Collaborators
  - Research - University of Virginia (Ravi), Google Research (Basu), University de Orleans (Bich, Christel Vrain), University de Lyon (Julien Velcin), University of Louvain (Siegfried Nijssen, Pierre Schaus),
  - Applications – Naval Medical Research Center, UC Davis Center for Imaging, Pennington Institute for Neuroscience, SoarTech
- Funding
  - 2019- National Science Foundation – “Explaining Unsupervised Learning – Combinatorial Optimization Approaches”
  - 2019- Google – “Combining Symbolic Reasoning and Deep Learning: A Constrained Optimization Formulation”
  - 2018 - Office of Naval Research – “Deep Graph Learning”
  - In the past NSF, OSD, Google, Yahoo, ONR