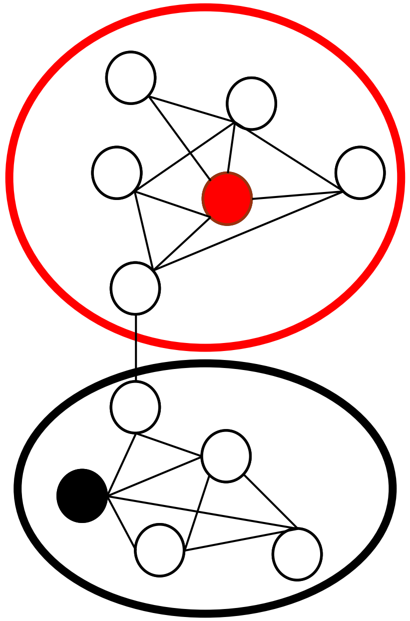
Graph-based Semi-supervised Learning: Realizing Pointwise Smoothness Probabilistically

Yuan Fang, University of Illinois at Urbana-Champaign
Kevin Chang, University of Illinois at Urbana-Champaign
Hady Lauw, Singapore Management University

06/24/2014 @ ICML

Graph-based semi-supervised learning

Data points: $X \in \mathcal{X} = \{x_1, \ldots, x_n\}$
Labels: $Y \in \mathcal{Y}$, the label of X



Labeled data

+

Structures in unlabeled data

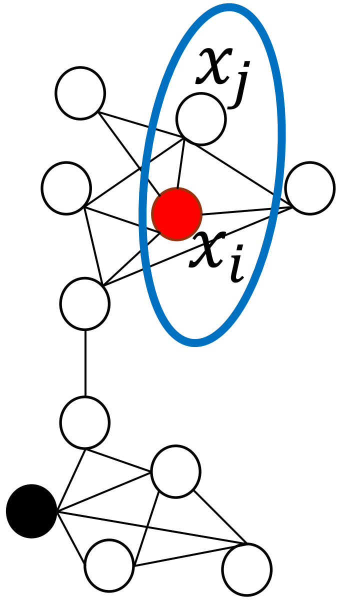
$$W_{ij} = \begin{cases} \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right) & i \neq j \\ 0 & i = j \end{cases}$$

Affinity: *pairwise geodesic distance*

Central notion in SSL: smoothness

“ Two points x_i, x_j are close

Their respective labels y_i, y_j are likely to be the same



Current graph-based smoothness does not precisely realize the nature of smoothness

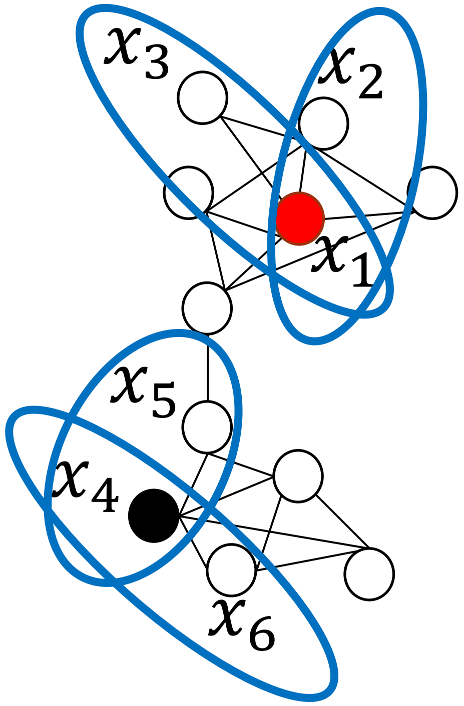
Realizing **Pointwise Nature**

through **Probabilistic Modeling**

Goal 1: Pointwise nature of smoothness

Inherently a property occurring
"everywhere"
on every point

Relating the behaviour of *each* point to that of its close points



Precisely expressing pointwise smoothness

If two points are close, their labels are likely to be the same.

Vague!

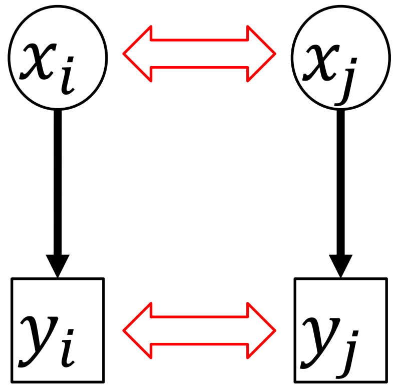
(Rigollet, 2007; Lafferty+, 2007)

P1. How to judge if two points are close?

(Data Closeness)

P2. Meaning of labels "likely to be the same"?

(Label Coupling)



Existing graph-based methods only express aggregate, not pointwise, smoothness

$$\min \sum_{ij} W_{ij}(y_i - y_j)^2$$

aggregate pairwise differences

- **"Average"** smooth across the graph
- **Not pointwise** formulation (P1 & P2)

Not an explicit realization of smoothness!

Goal 2: Probabilistic model of smoothness

P1. How to judge if two points are close?

How close is *sufficiently* close? → Deterministic binary decision ✗

Probabilistic formulation ✓

P2. Meaning of labels "likely to be the same"?

Deterministic ✗ Probabilistic ✓

Probabilistic Modelling: $P(Y|X)$, or $P(X|Y)$ and $P(Y)$

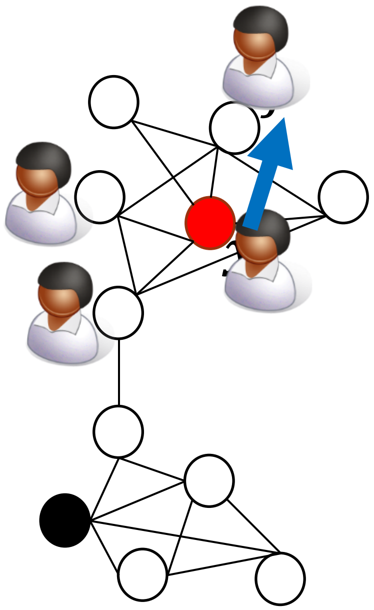
Our Proposal: Probabilistic Graph-based Pointwise Smoothness (PGP)

- **P1. Data Closeness**
  - Capture closeness distribution of two points X, X' on graph
- **P2. Label Coupling**
  - Couple the label distribution of two close points X, X'

P1. Data Closeness:
Random walk on graph

- Consider a random walk on the graph $\{V_t : t = 0,1,\ldots\}$
  - A traversal from x_i to x_j $(V_t = x_i, V_{t+1} = x_j)$

a traversal from X to X' → "connects" X, X' as close points



- The event that x_i, x_j close: $(X = x_i, X' = x_j)$
  - Distribution of traversal from x_i to x_j in the long run $(t \to \infty)$
  - (X, X') is a pair of limiting random variables:

$$(V_t, V_{t+1}) \xrightarrow{d} (X, X')$$

P2. Label Coupling:

Statistical indistinguishability

X is x_i itself
($X = x_i$)

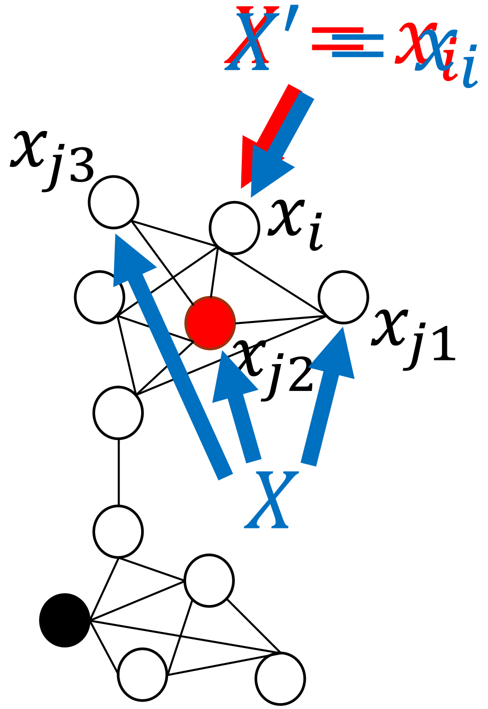
or

X is close to x_i
($X' = x_i$)

Y of X distributes similarly

$P(Y|X = x_i)$
$P(Y|X, X' = x_i)$ are alike

Statistical Indistinguishability



P2. Label Coupling:

Statistical indistinguishability

$$p(Y|X = x_i) \qquad \bigcirc \qquad p(Y|X, X' = x_i)$$

⇩

α-statistical indistinguishable

Intuition: indistinguishability will slowly fade along a "chain" of close points

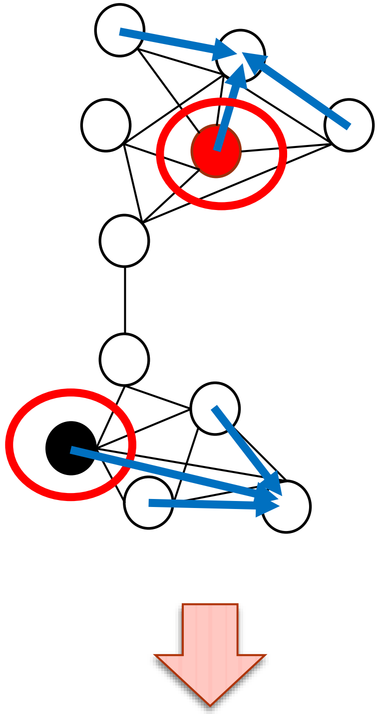
Constraint-based solution for $P(X|Y=y)$

Labelled data:

$$
\begin{aligned}
& p(X=x_i|Y=y) \\
&= p(Y=y|X=x_i)\,p(X=x_i)\,/p(Y=y) \\
&\propto p(Y=y|X=x_i)\,Z_i
\end{aligned}
$$

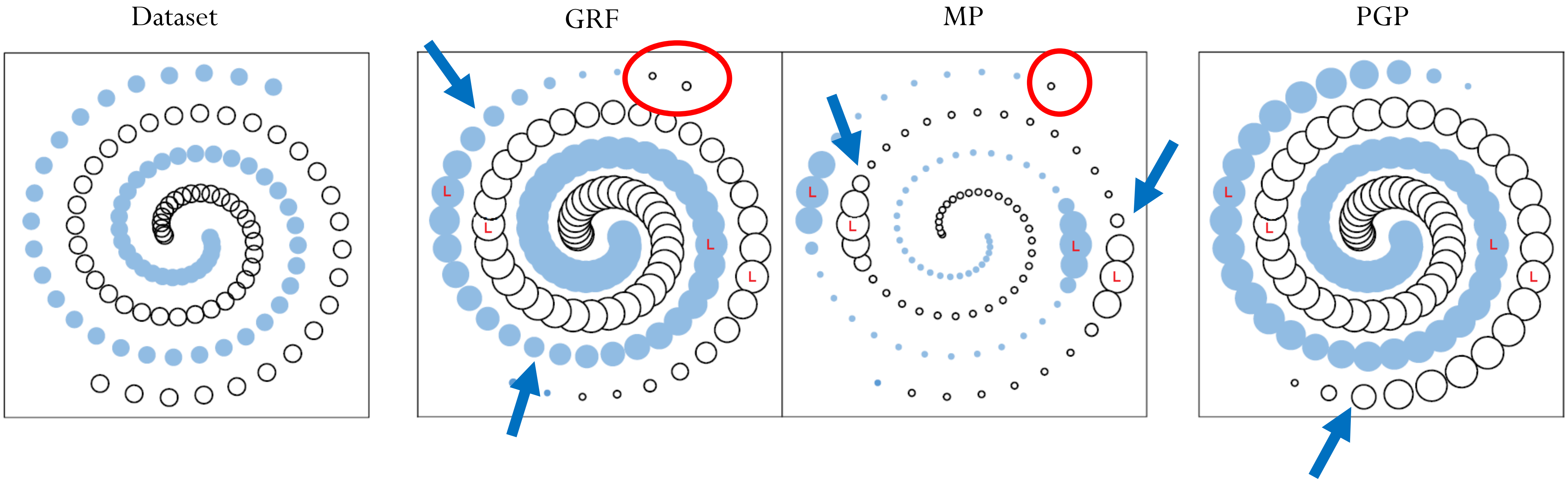
Unlabelled data (smoothness P1 & P2):

$$
\begin{aligned}
p(X=x_i|Y=y) &\overset{1}{=} (1-\alpha)p(X,X'=x_i|Y=y) \\
&\overset{2}{=} (1-\alpha)\textstyle\sum_j p(X=x_j,X'=x_i|Y=y) \\
&\overset{3}{=} (1-\alpha)\textstyle\sum_j p(X'=x_i|X=x_j,Y=y)\,p(X=x_j|Y=y) \\
&\overset{4}{=} (1-\alpha)\textstyle\sum_j p(X'=x_i|X=x_j)\,p(X=x_j|Y=y) \\
&\overset{5}{=} (1-\alpha)\textstyle\sum_j W_{ji}/Z_j \cdot p(X=x_j|Y=y)
\end{aligned}
$$



unique solution
to $p(X|Y=y)$

Experiment 1: PGP is smooth “everywhere"



L: labeled point

Size: magnitude of decision function

Color: predicted class

Non-smooth area

Misclassification

Experiment 2: PGP is consistently the best

	$\lvert\mathcal{L}\rvert = 10$					
	Digit1	Text	ISOLET	Cancer	USPS	Yeast
GRF	.894	.451	.627	.871	.638	.510
LSVM	.833	.428	.719	.886	**.698**	.562
GGS	.855	.567	.677	.867	.666	.540
MP	**.901**	.558	.692	.898	**.713**	.574
PARW	.881	**.587**	.721	.893	**.706**	.575
PGP	**.910**	**.592**	**.734**	**.910**	**.704**	**.593**

Conclusion

- Smoothness is pointwise in nature
- Probabilistic modelling of smoothness