Semi-Supervised Learning on Data Streams via Temporal Label Propagation

Tal Wagner — Massachusetts Institute of Technology

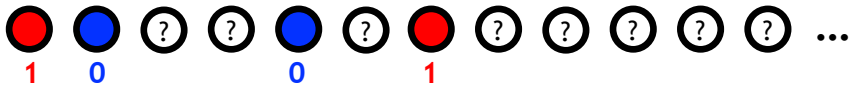
Joint work with:

Sudipto Guha — amazon

Shiva Kasiviswanathan — amazon

Nina Mishra — amazon

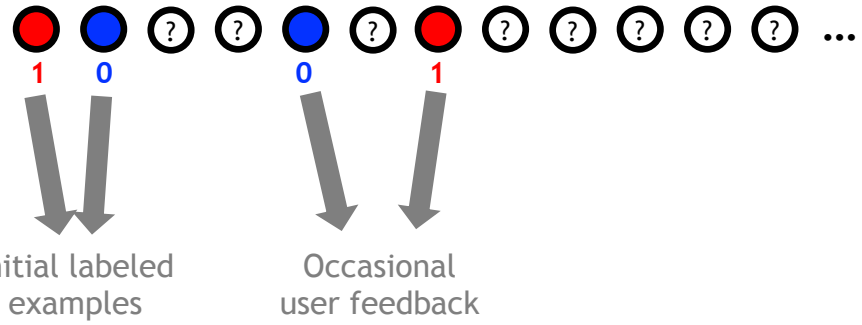
Semi-Supervised Learning on a Stream



Problem:

- Data points arrive on a stream
- Few are labeled
- Most are unlabeled
- Task: Label points on-the-fly

Semi-Supervised Learning on a Stream



Goals:

- **Time**: Label points quickly
- **Space**: Stream too large to fully store
- Learn from unlabeled data

Background:

Offline Semi-Supervised Learning

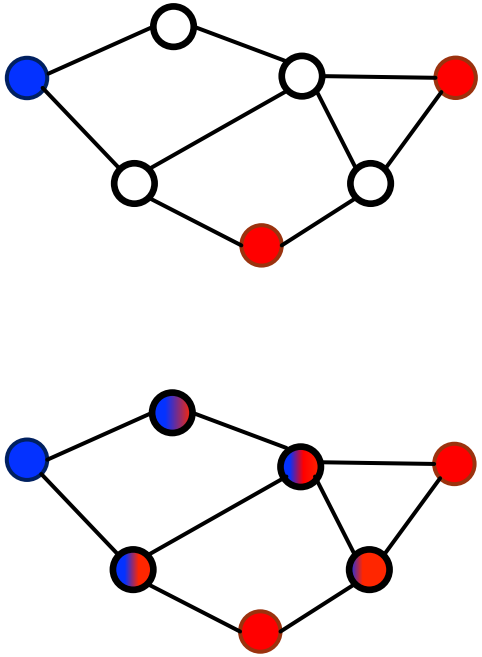
Offline Semi-Supervised Learning:

Label Propagation

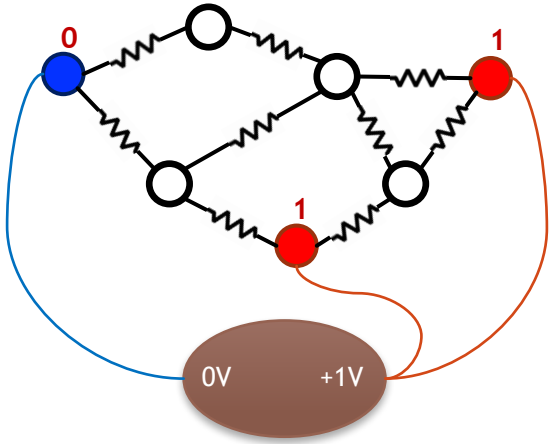
[Zhu, Ghahramani, Lafferty ICML'03]

[ICML 10-Year Classic Paper Prize 2013]

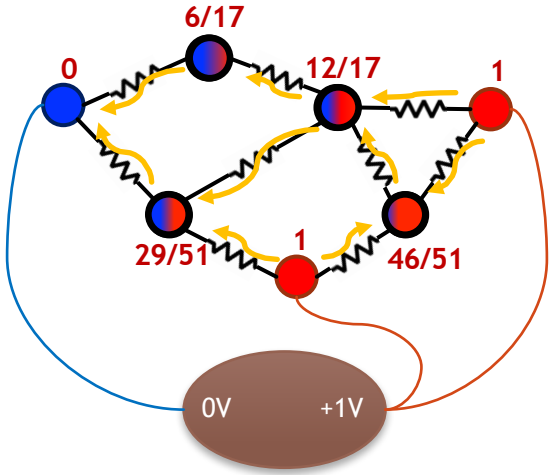
- Construct graph on data points
- Propagate labels on graph



Label Propagation: Electric Intuition

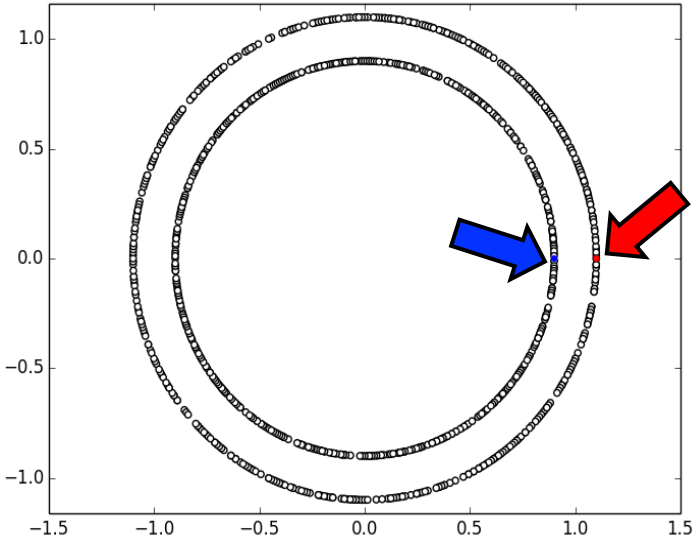


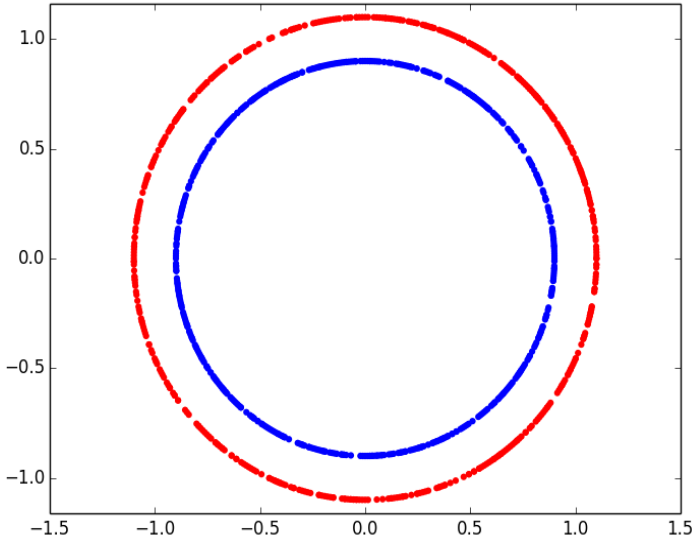
Label Propagation: Electric Intuition



$f: V \to [0,1]$ computed by $\min_f \sum_{(x,y)\in E} w_{x,y}\big(f(x) - f(y)\big)^2$ **s.t.** $f(\bigcirc) = 0,\ f(\bigcirc) = 1$

Label Propagation: In Action (I)





Label Propagation: In Action (II)

[Levin, Lischinski, Weiss SIGGRAPH'04]

Input

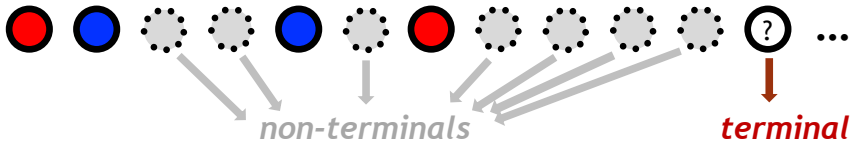
Labeled examples

Output

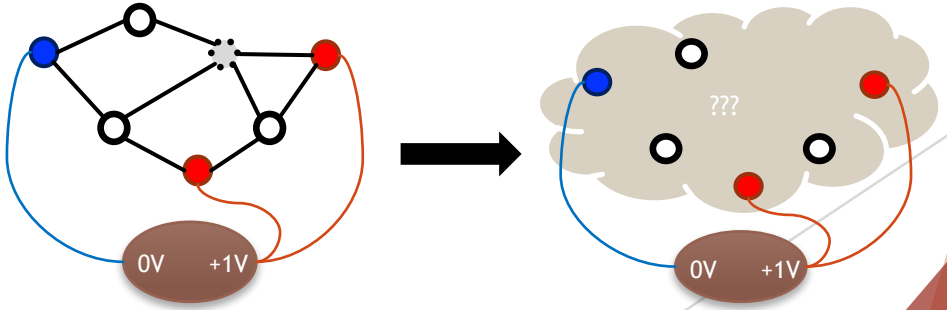
Handling Data Streams:

From Offline to Streaming Label Propagation

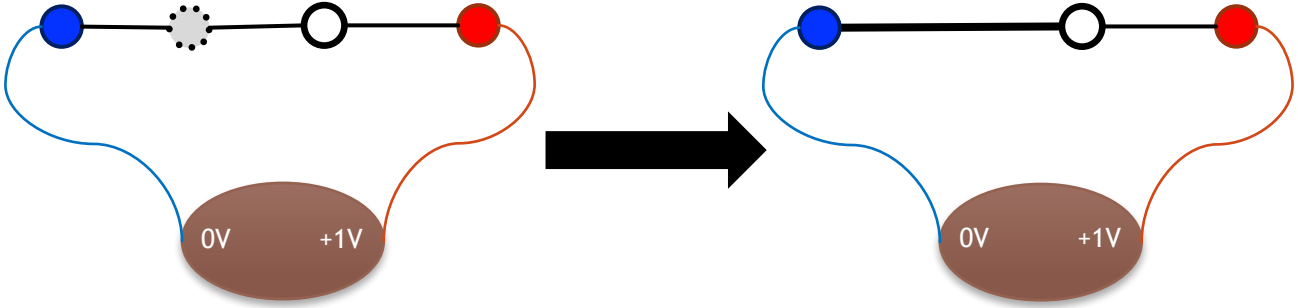
Semi-Supervised Learning on a Stream



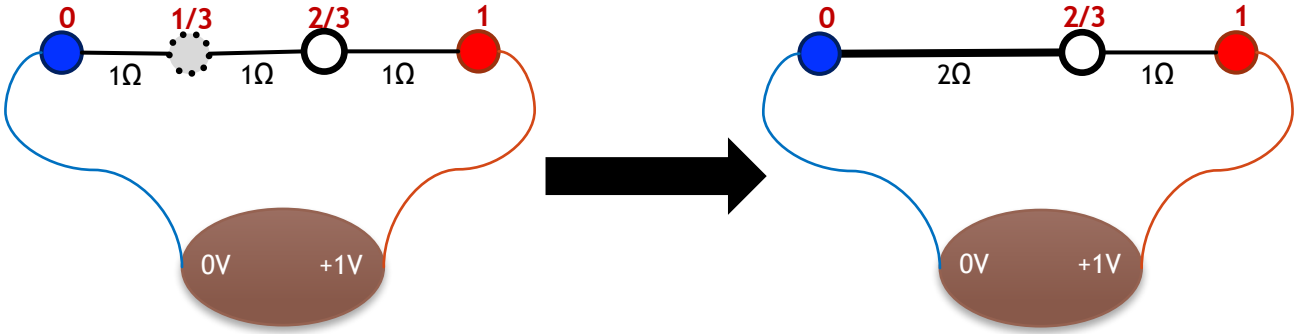
- Observe: We only need to label the new point
  - Offline algorithm labels all nodes at once – redundant work
- However: Terminal label depends on all non-terminals
- **Can we get rid of non-terminals?**



Graph Reduction

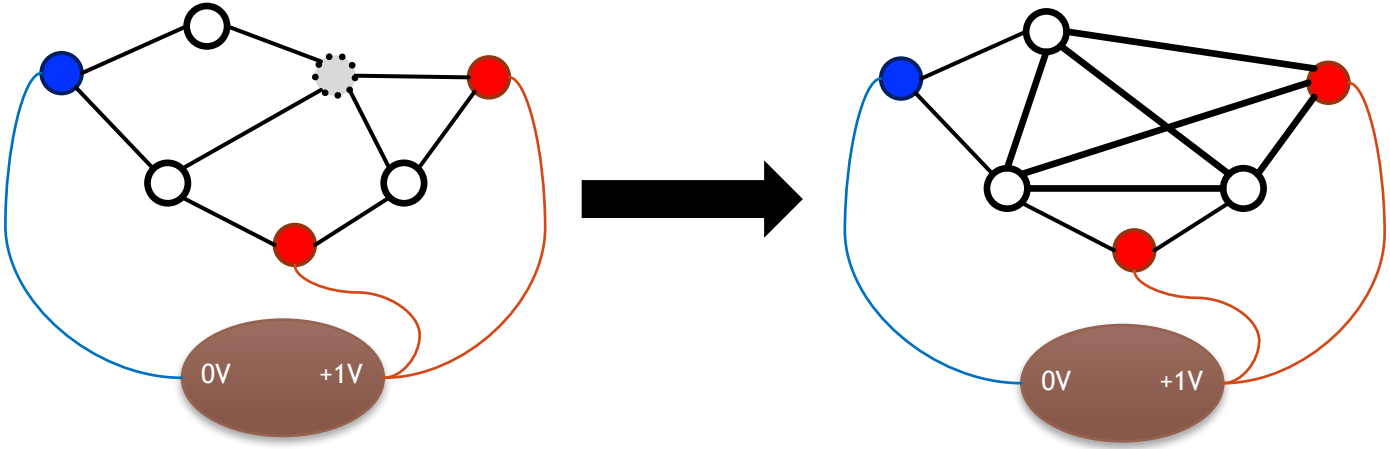


Graph Reduction



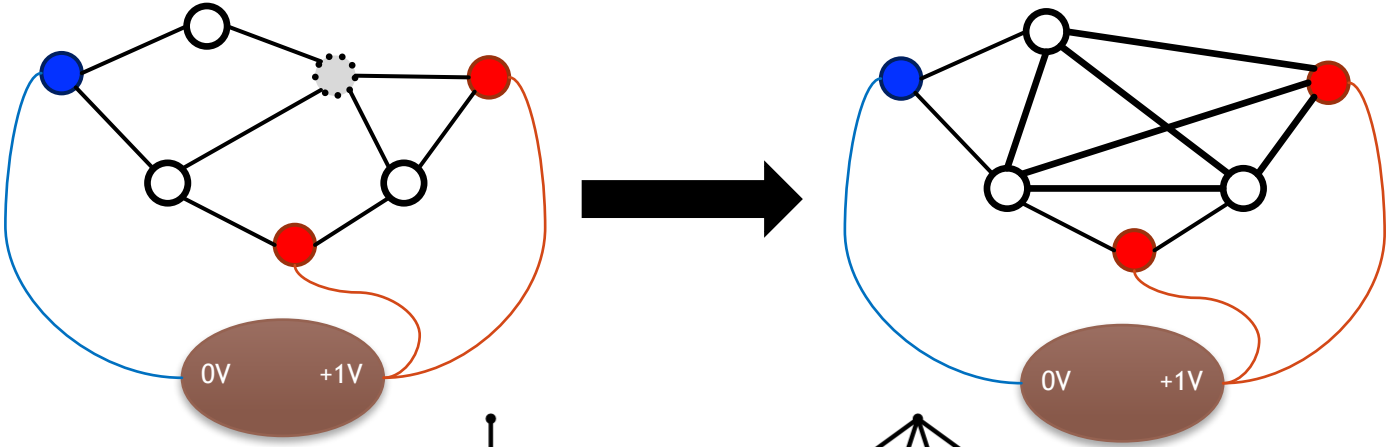
Graph Reduction:

Star-Mesh Transform

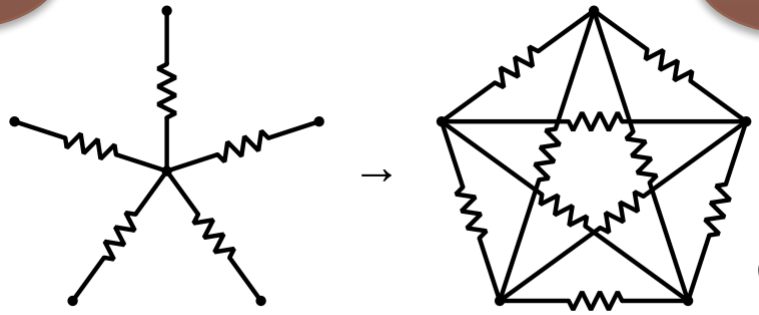


Graph Reduction:

Star-Mesh Transform



"Star-mesh transform" [Rosen 1924]

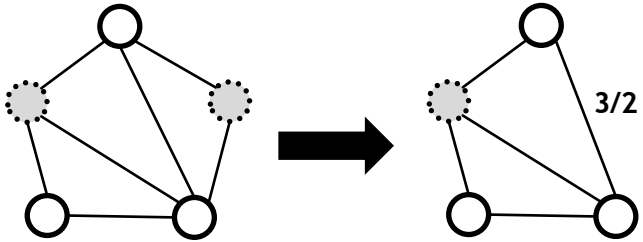


"star" "mesh"

(from Wikipedia)

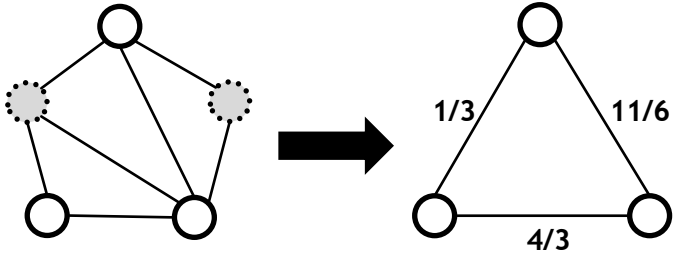
Graph Reduction:
Sequential Star-Mesh Transforms

- Graph has n nodes and τ terminals
- Star-mesh the $n - \tau$ non-terminals one by one



Graph Reduction:
Sequential Star-Mesh Transforms

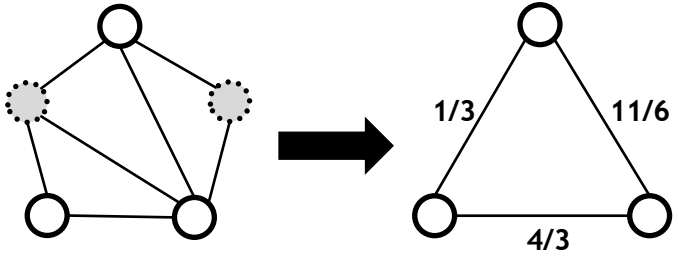
- Graph has n nodes and τ terminals
- Star-mesh the $n - \tau$ non-terminals one by one
  - Computes Schur complement of graph Laplacian matrix [Campbell 1922]



Graph Reduction:

Sequential Star-Mesh Transforms

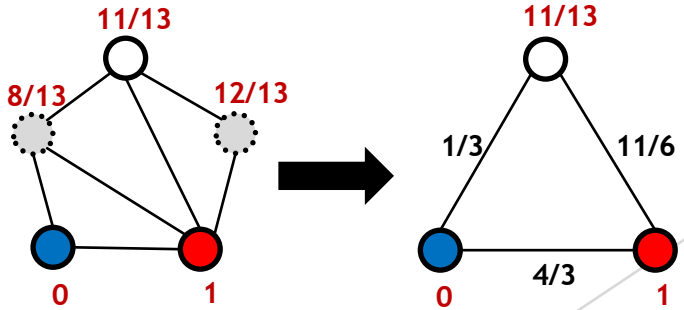
- Graph has n nodes and τ terminals
- Star-mesh the $n - \tau$ non-terminals one by one
  - Computes Schur complement of graph Laplacian matrix [Campbell 1922]



- **Compression:**

  From $\Omega(n \log n)$ to $O(\tau^2 \log n)$ bits

- **Theorem:** Label Propagation is preserved on terminals.

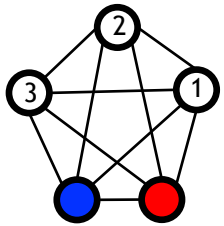


Our Algorithm:

Temporal Label Propagation (TLP)

Temporal Label Propagation

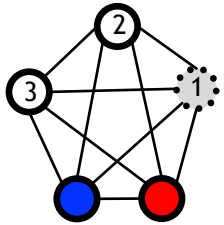
- Pick small integer τ
- Store τ recent unlabeled points



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

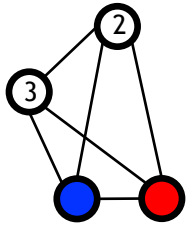
- On point arrival:
  - **Star-mesh** out oldest point



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

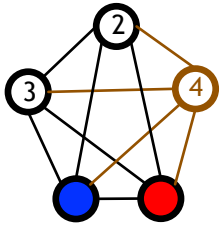
- On point arrival:
  - **Star-mesh** out oldest point



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

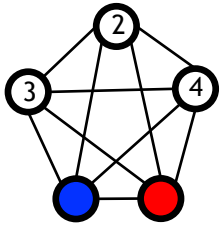
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

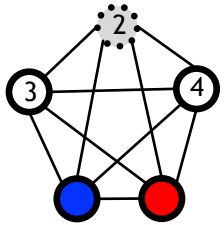
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point
  - **Propagate** labels



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

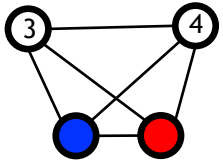
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point
  - **Propagate** labels



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

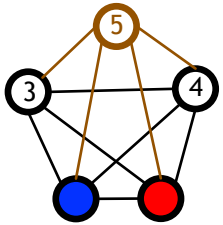
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point
  - **Propagate** labels



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

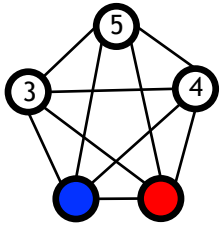
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point
  - **Propagate** labels



Temporal Label Propagation

- Pick small integer τ
- Store τ recent unlabeled points

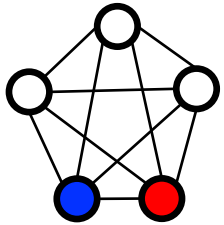
- On point arrival:
  - **Star-mesh** out oldest point
  - **Insert** new point
  - **Propagate** labels



Temporal Label Propagation

Guarantees for time step n:

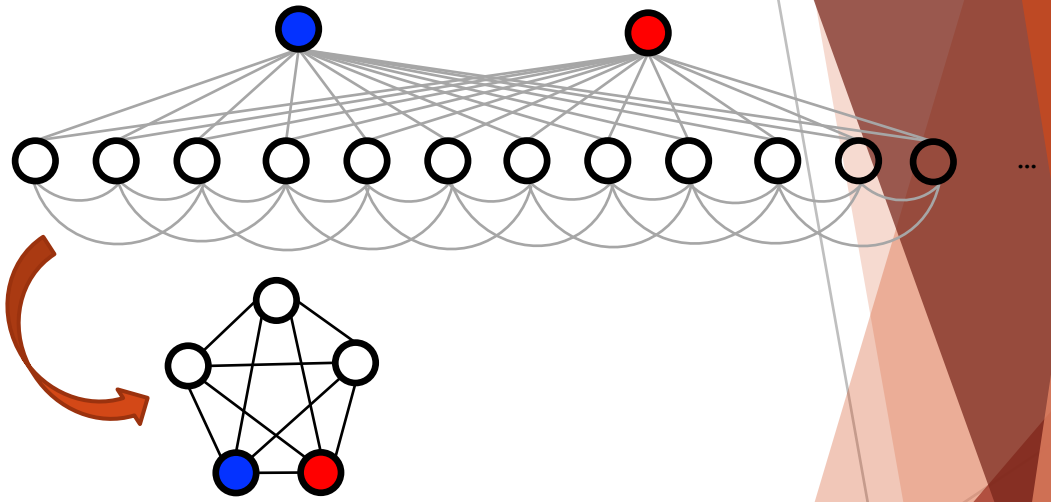
- Processing time: $\widetilde{O}(\tau^3)$
- Storage space: $\widetilde{O}(\tau^2 \log n)$



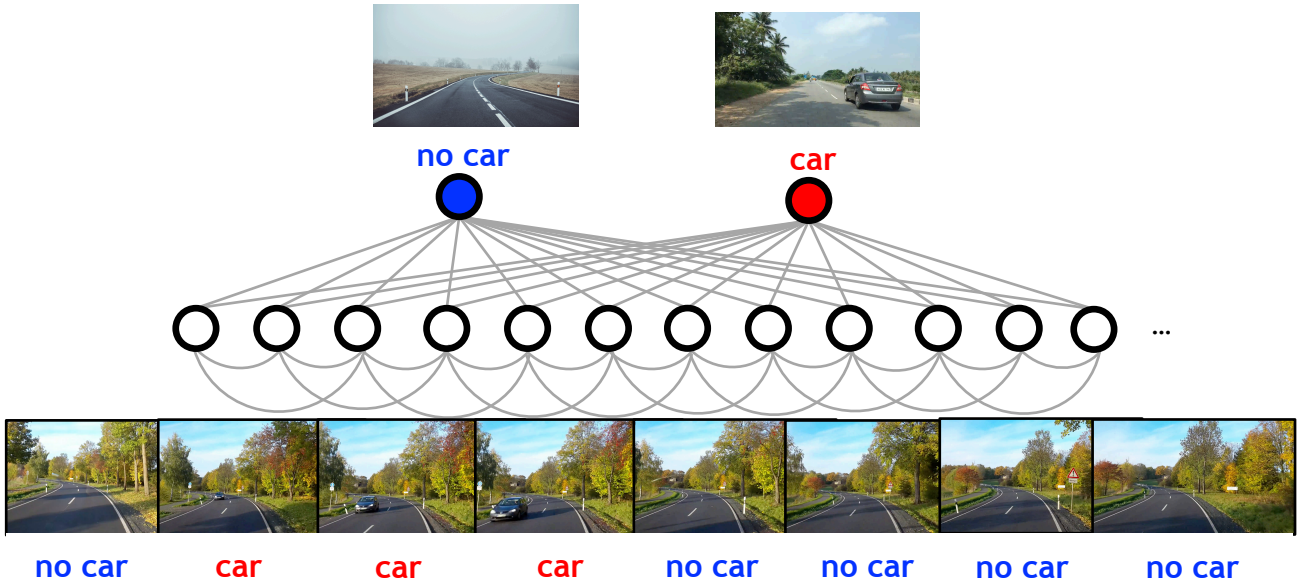
Temporal Label Propagation

Guarantees for time step n:

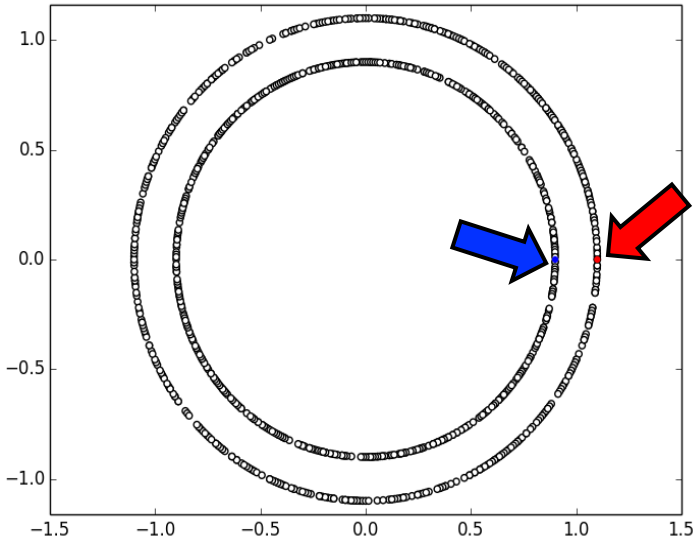
- Processing time: $\widetilde{O}(\tau^3)$
- Storage space: $\widetilde{O}(\tau^2 \log n)$
- Labels propagate on a graph containing the entire stream so far

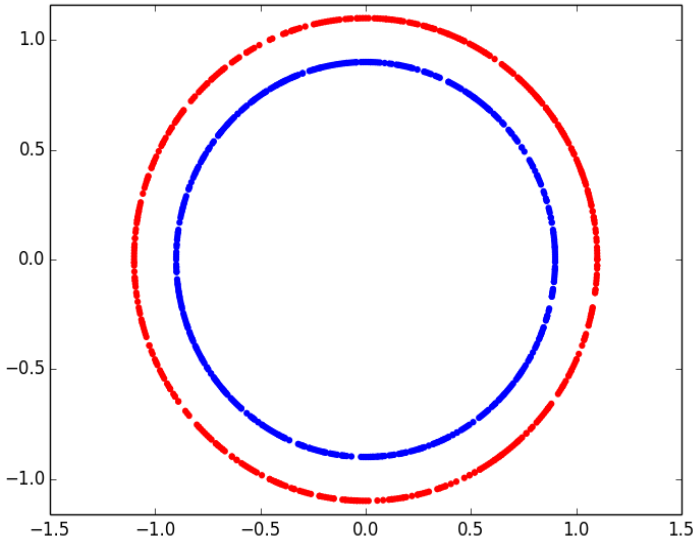


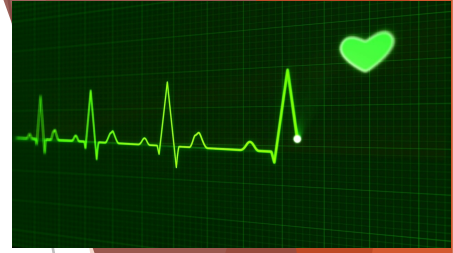
Temporal Vicinity Graph



Demo: Two Rings in Angular Order

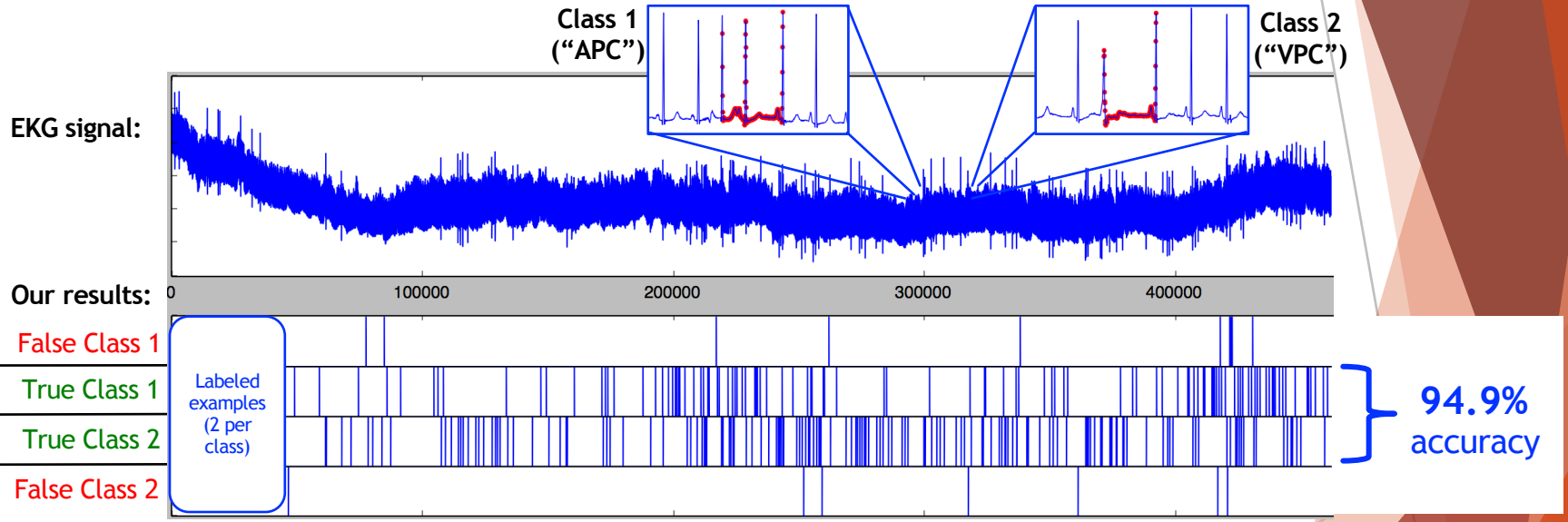






Real Data:

Classifying Irregular Heartbeats on EKG



- Data from **physionet.org**
- Class 1: Atrial premature contraction (APC)
- Class 2: Ventricular premature contraction (VPC)
- Normal heartbeats are ignored
- Overall accuracy: **94.9%** with $\tau = 5$ and **2** labeled examples per class

Conclusion

- **Principled and practical approach to streaming SSL**
  - Rich, robust, well-studied mathematical framework
  - Simple to implement
- **Future directions:**
  - **More applications of compression for SSL**
    - Distributed networks, memory-limited devices (**edge**, **GPU**, …)
  - **Extensions and variants of Label Propagation on streams**
    - Support **regularization**, **noisy labels**, **interpretability**, …
  - **Trade approximation / randomization for even better performance**
    - Our algorithm is **exact** and **deterministic**

Thank you

Questions?