1st BRICS Countries Congress (BRICS-CCI) and 11th Brazilian Congress (CBIC) on Computational Intelligence

Combined Active and Semi-Supervised Learning using Particle Walking Temporal Dynamics

Fabricio Breve fabricio@rc.unesp.br

Department of Statistics, Applied Mathematics and Computation (DEMAC), Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (UNESP), Rio Claro, SP, Brazil

Outline

- Active Learning and Semi-Supervised Learning
- The Proposed Method
- Computer Simulations
- Conclusions

Active Learning

- Learner is able to interactively query an human specialist (or some other information source) to obtain the labels of selected data points
- **Key idea:** greater accuracy with fewer labeled data points

[4] B. Settles, “Active learning,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, no. 1, pp. 1–114, 2012.
[5] F. Olsson, “A literature survey of active machine learning in the context of natural language processing,” Swedish Institute of Computer Science, Box 1263, SE-164 29 Kista, Sweden, Tech. Rep. T2009:06, April 2009.

Semi-Supervised Learning

- Learns from both labeled and unlabeled data items.
  - Focus on problems where there are lots of easily acquired unlabeled data, but the labeling process is expensive, time consuming, and often requiring the work of human specialists.

[1] X. Zhu, “Semi-supervised learning literature survey,” Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2005.
[2] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press, 2006.
[3] S. Abney, *Semisupervised Learning for Computational Linguistics*. CRC Press, 2008.

Semi-Supervised Learning and Active Learning comparison

Semi-Supervised Learning

- Exploits what the learner thinks it knows about the unlabeled data
- Most confident labeled data used to retrain algorithm *(self-learning methods)*
- Relies on committee agreements *(co-training methods)*

Active Learning

- Attempt to explore unknown aspects of the data
- Less confident labeled data have their labels queried *(uncertainty sampling methods)*
- Query according to committee disagreements *(query by committee methods)*

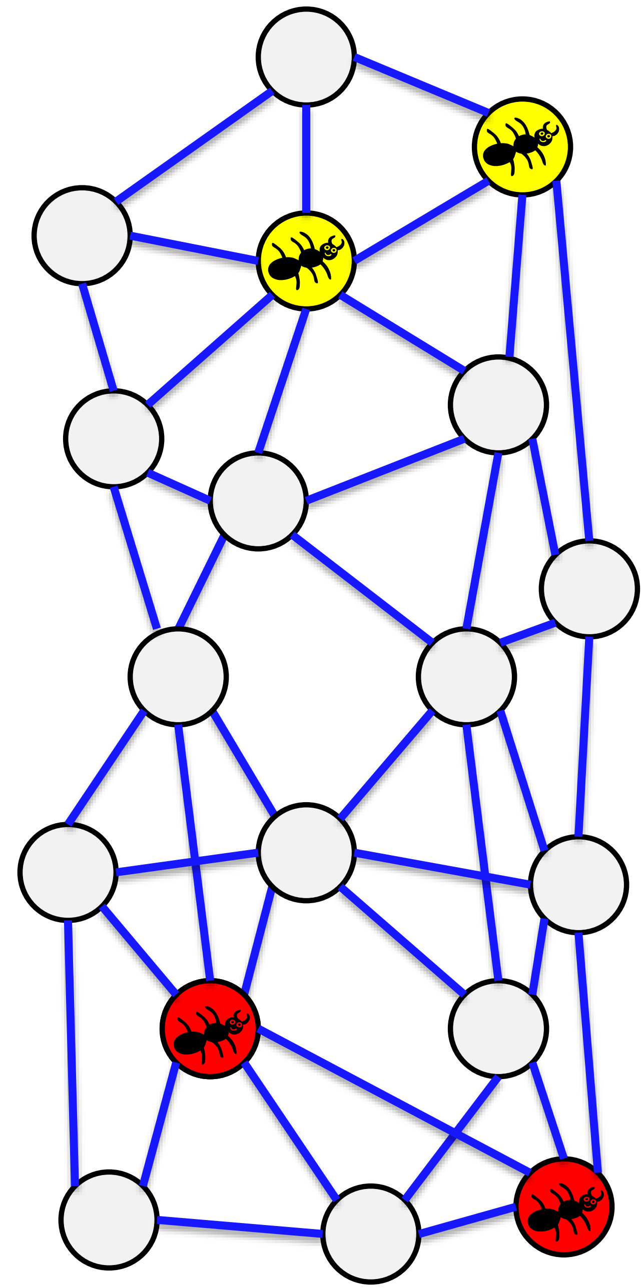
[4] B. Settles, "Active learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 6, no. 1, pp. 1–114, 2012.
[5] F. Olsson, "A literature survey of active machine learning in the context of natural language processing," Swedish Institute of Computer Science, Box 1263, SE-164 29 Kista, Sweden, Tech. Rep. T2009:06, April 2009.

Proposed Method

- Semi-Supervised Learning and Active Learning combined into a new nature-inspired method
  - Particles competition and cooperation in networks combined into an unique schema
    - Cooperation:
      - Particles from the same class (team) walk in the network cooperatively, propagating their labels.
      - **Goal**: Dominate as many nodes as possible.
    - Competition:
      - Particles from different classes (teams) compete against each other
      - **Goal**: Avoid invasion by other class particles in their territory

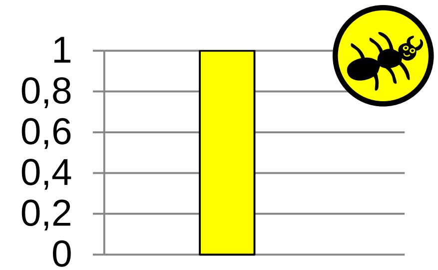
Initial Configuration

- An undirected network is generated from data by connecting each node to its k-nearest neighbors
- A particle is generated for each labeled node of the network
- Particles initial position are set to their corresponding nodes
- Particles with same label play for the same team

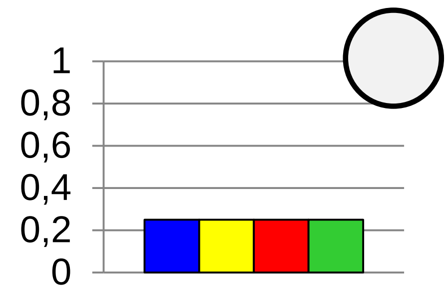


Initial Configuration

- Nodes have a domination vector
  - Labeled nodes have ownership set to their respective teams (classes).
  - Unlabeled nodes have levels set equally for each team



Ex: [0.00 1.00 0.00 0.00]
(4 classes, node labeled as class B)

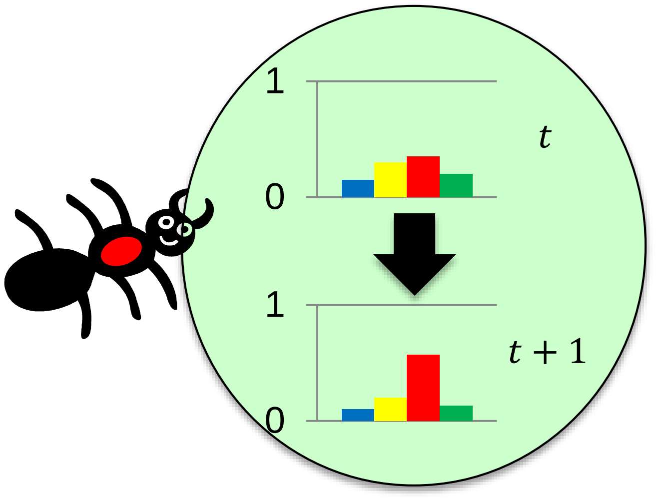


Ex: [0.25 0.25 0.25 0.25]
(4 classes, unlabeled node)

$$v_i^{\omega_\ell} = \begin{cases} 1 & \text{if } y_i = \ell \\ 0 & \text{if } y_i \neq \ell \text{ e } y_i \in L \\ {}^{1}/_{c} & \text{if } y_i = \emptyset \end{cases}$$

Node Dynamics

- When a particle selects a neighbor to visit:
  - It **decreases** the domination level of the **other** teams
  - It **increases** the domination level of its **own** team
  - Exception: **labeled** nodes domination levels are **fixed**

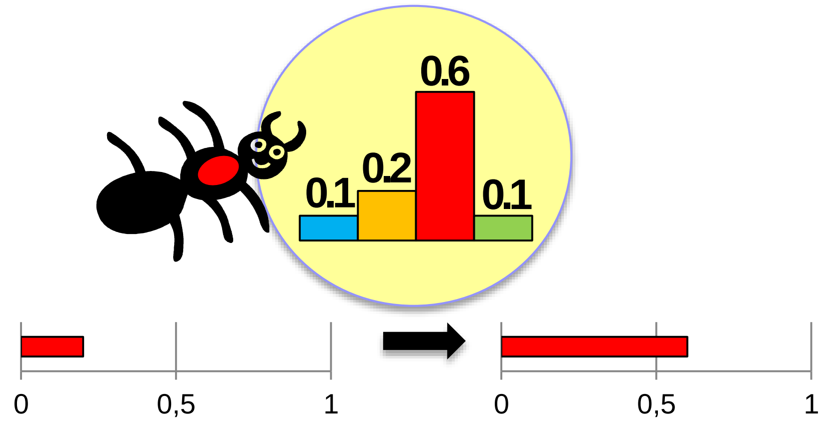


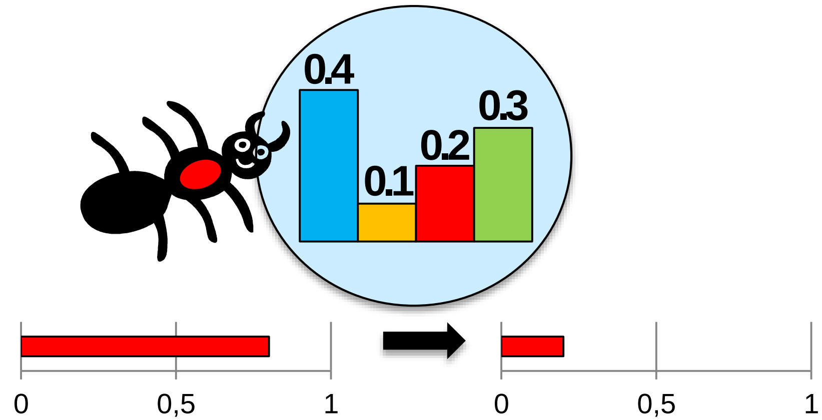
$$v_i^{\omega_\ell}(t+1) = \begin{cases} \max\left\{0, v_i^{\omega_\ell}(t) - \dfrac{0.1\,\rho_j^{\omega}(t)}{c-1}\right\} & \text{se } \ell \neq \rho_j^f \\ v_i^{\omega_\ell}(t) + \sum_{r\neq\ell} v_i^{\omega_r}(t) - v_i^{\omega_r}(t+1) & \text{se } \ell = \rho_j^f \end{cases}$$

Particle Dynamics

- A particle gets:
  - **Stronger** when it selects a node being dominated by **its own** team
  - **Weaker** when it selects a node being dominated by **another** team

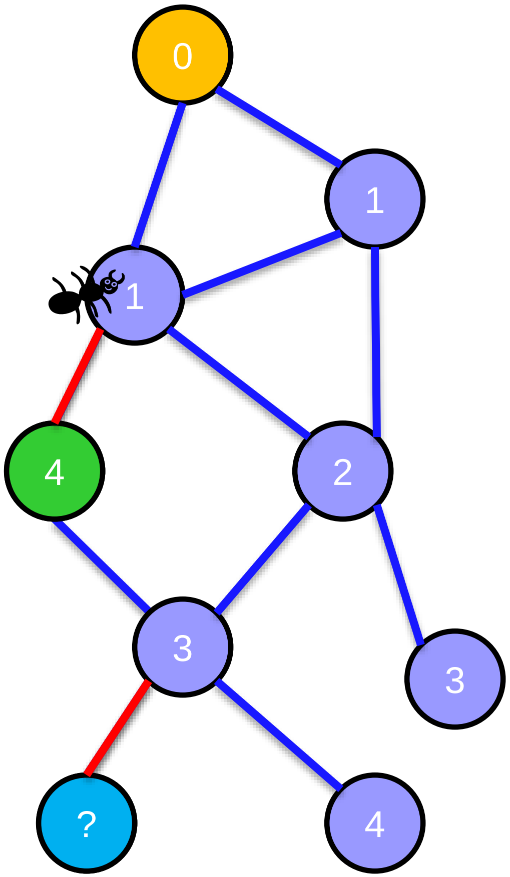
$$\rho_j^{\omega}(t) = v_i^{\omega_\ell}(t)$$





Distance Table

- Each particle has its distance table.
- Keep the particle aware of how far it is from the closest labeled node of its team (class).
  - Prevents the particle from losing all its strength when walking into enemies neighborhoods.
  - Keeps the particle around to protect its own neighborhood.
- Updated dynamically with local information.
  - No prior calculation.

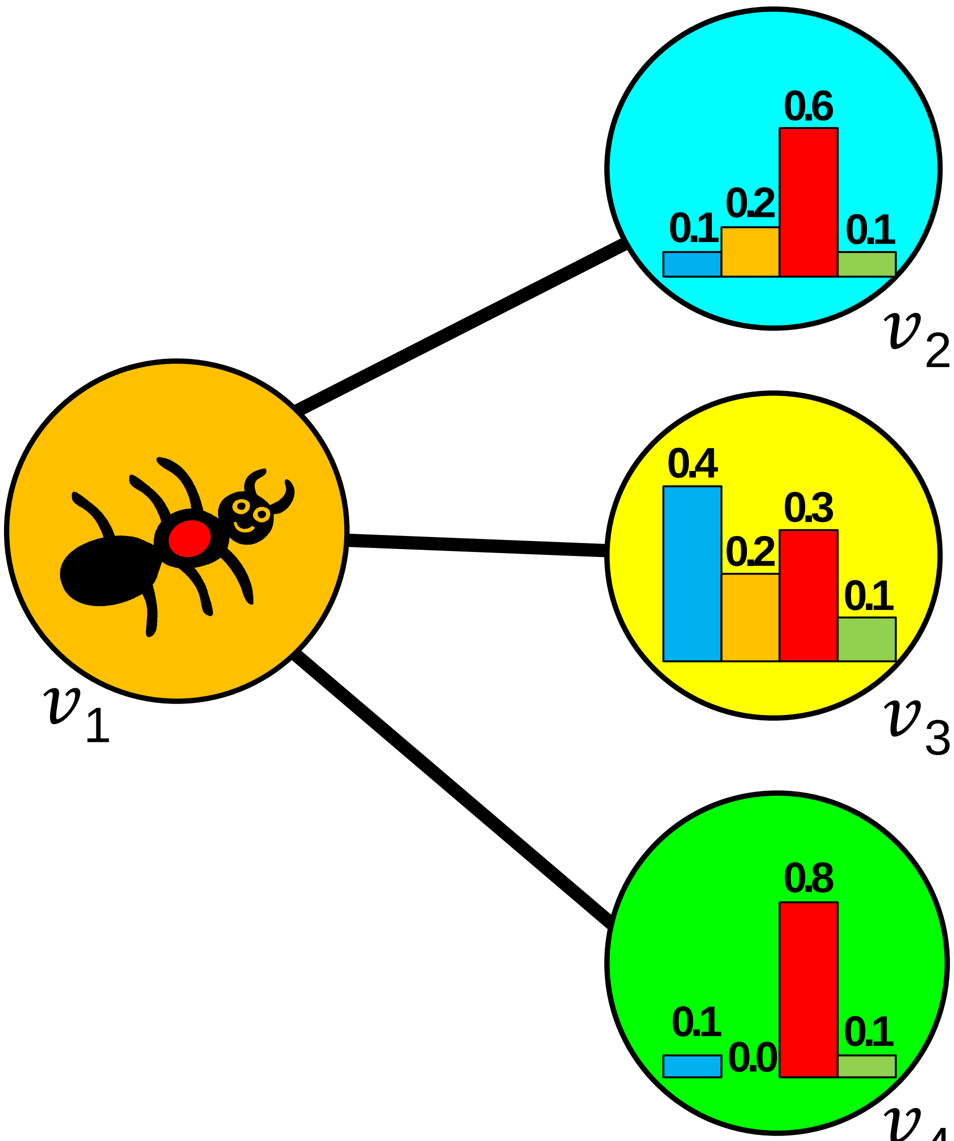


$$\rho_j^{d_k}(t+1) = \begin{cases} \rho_j^{d_i}(t)+1 & \text{se } \rho_j^{d_i}(t)+1 < \rho_j^{d_k}(t) \\ \rho_j^{d_k}(t) & \text{otherwise} \end{cases}$$

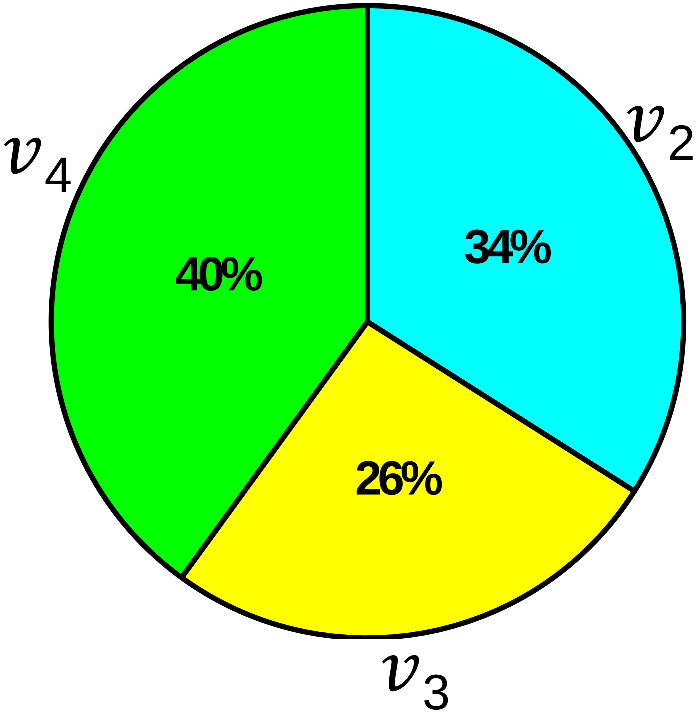
Particles Walk

- Random-greedy walk
  - Each particles randomly chooses a neighbor to visit at each iteration
  - Probabilities of being chosen are higher to neighbors which are:
    - Already dominated by the particle team.
    - Closer to particle initial node.

$$p\left(v_i|\rho_j\right) = \frac{W_{qi}}{2\sum_{\mu=1}^{n} W_{q\mu}} + \frac{W_{qi} v_i^{\omega_\ell}\left(1+\rho_j^{d_i}\right)^{-2}}{2\sum_{\mu=1}^{n} W_{q\mu}\, v_\mu^{\omega_\ell}\left(1+\rho_j^{d_\mu}\right)^{-2}}$$

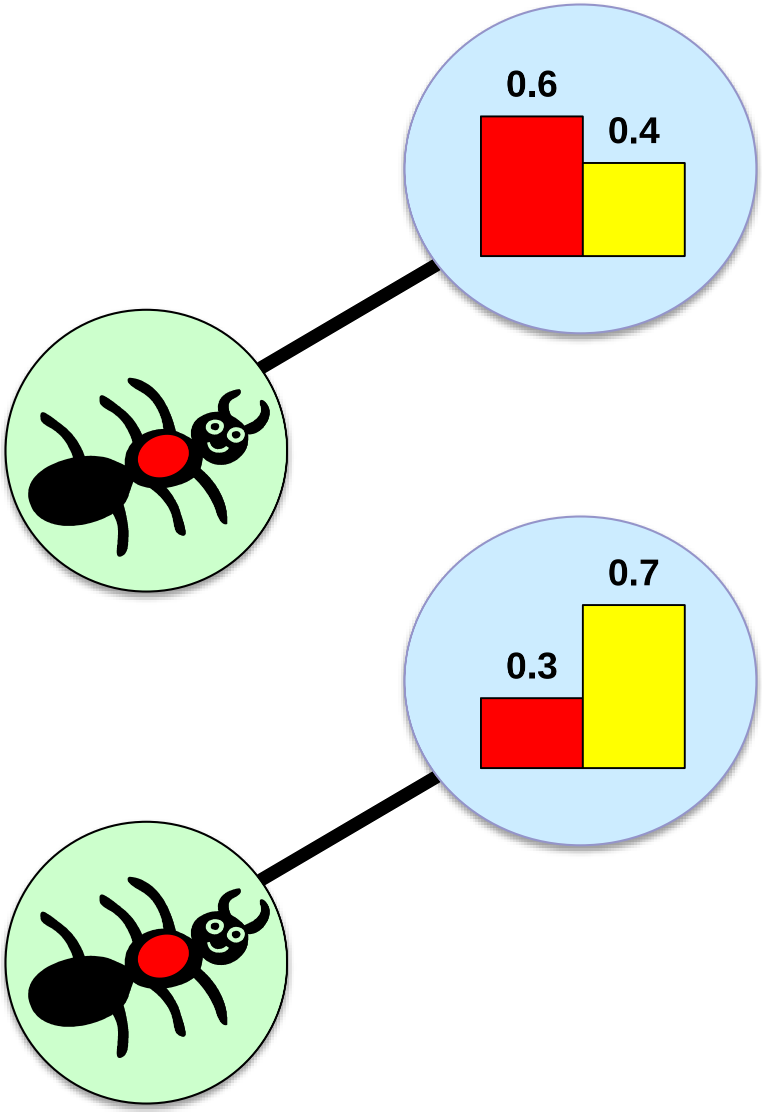


Moving Probabilities



Particles Walk

- Shocks
  - A particle really visits the selected node only if the domination level of its team is higher than others;
  - Otherwise, a shock happens and the particle stays at the current node until next iteration.



Label Query

- When the nodes domination levels reach a fair level of stability, the system chooses a unlabeled node and queries its label.
  - A new particle is created to this new labeled node.
  - The iterations resume until stability is reached again, then a new node will be chosen.
  - The process is repeated until the defined amount of labeled nodes is reached.

Query Rule

- Two versions of the algorithm:
  - **ASL-PCC A**
  - **ASL-PCC B**
- They use different rules to select which node will be queried.

ASL-PCC **A**

- Uses temporal node domination information to select the unlabeled node which had more dispute over time.
  - The node the algorithm has less confidence on the label it is currently assigning.

$$q(t) = \arg\max_{i, y = \emptyset} u_i(t)$$

$$u_i(t) = \frac{v_i^{\lambda_{\ell **}}(t)}{v_i^{\lambda_{\ell *}}(t)}$$

$$v_i^{\lambda_{\ell *}}(t) = \arg\max_{\ell} v_i^{\lambda_\ell}(t)$$

$$v_i^{\lambda_{\ell **}}(t) = \arg\max_{\ell, \ell \neq v_i^{\lambda_{\ell *}}(t)} v_i^{\lambda_\ell}(t)$$

AL-PCC **B**

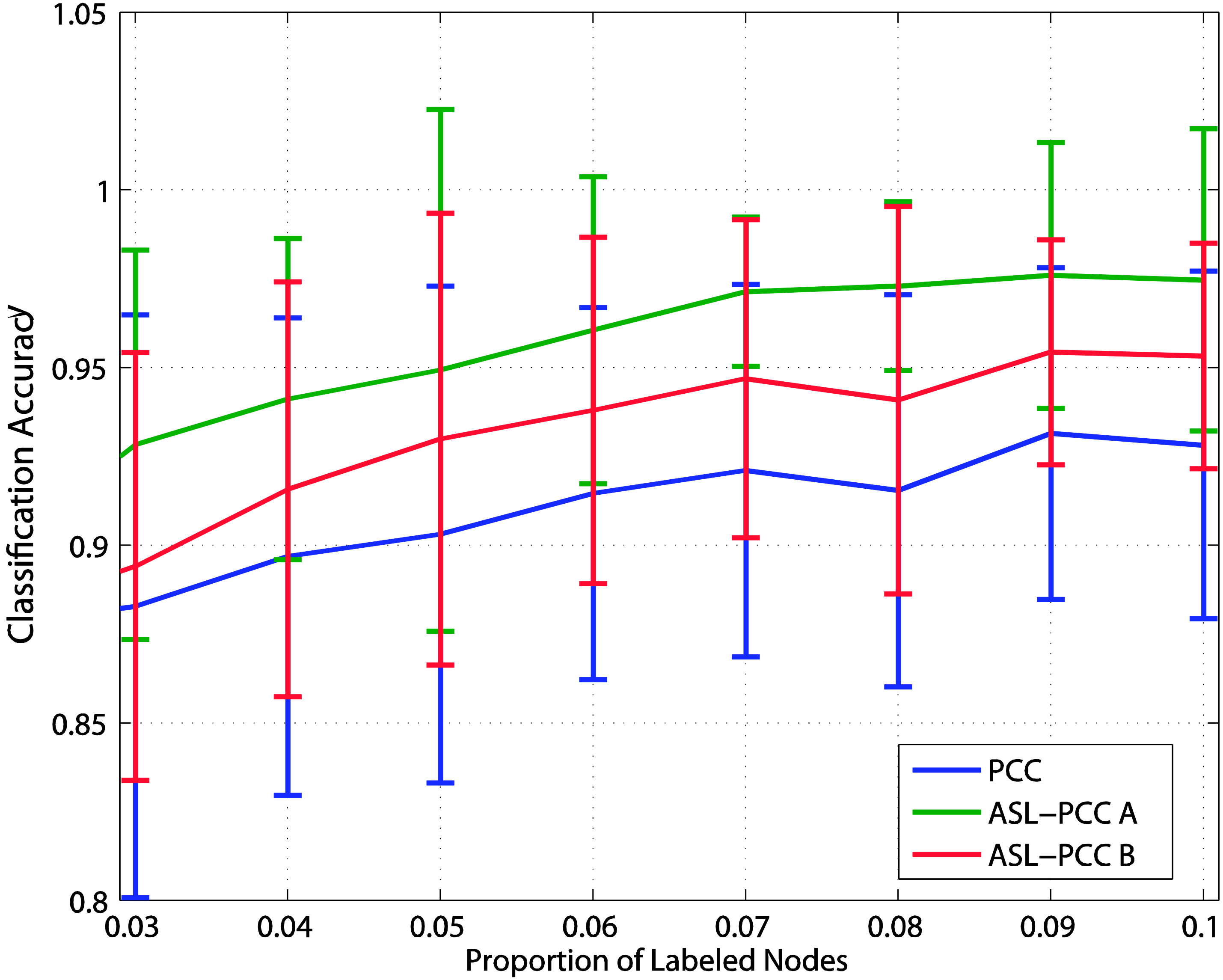
- Chooses the unlabeled node which is currently more far away from any labeled node.
  - According to particles dynamic distance tables.

$$s_i(t) = \min_j \rho_j^{d_i}(t)$$

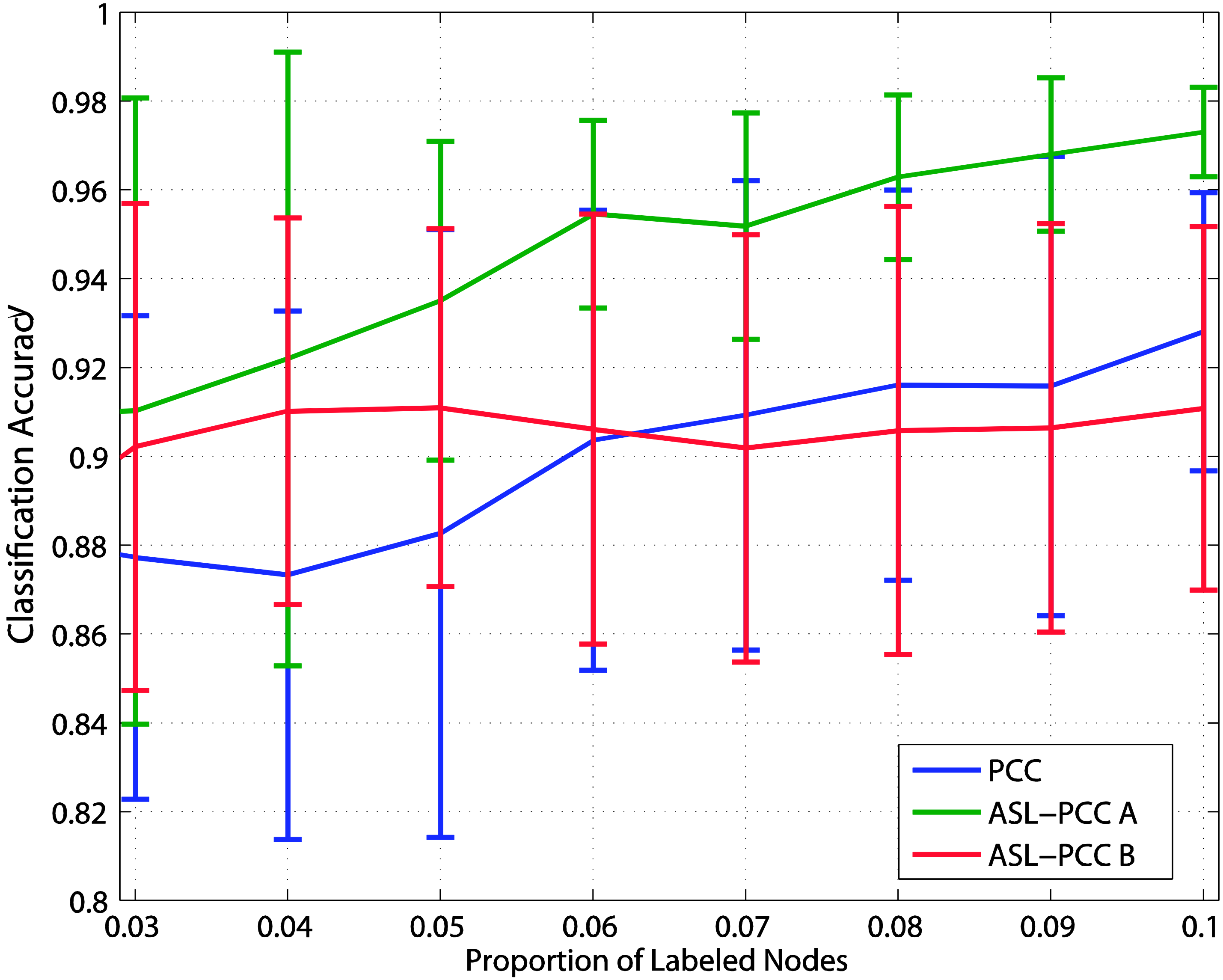
$$q(t) = \arg\max_{i, y=\emptyset} s_i(t)$$

Computer Simulations

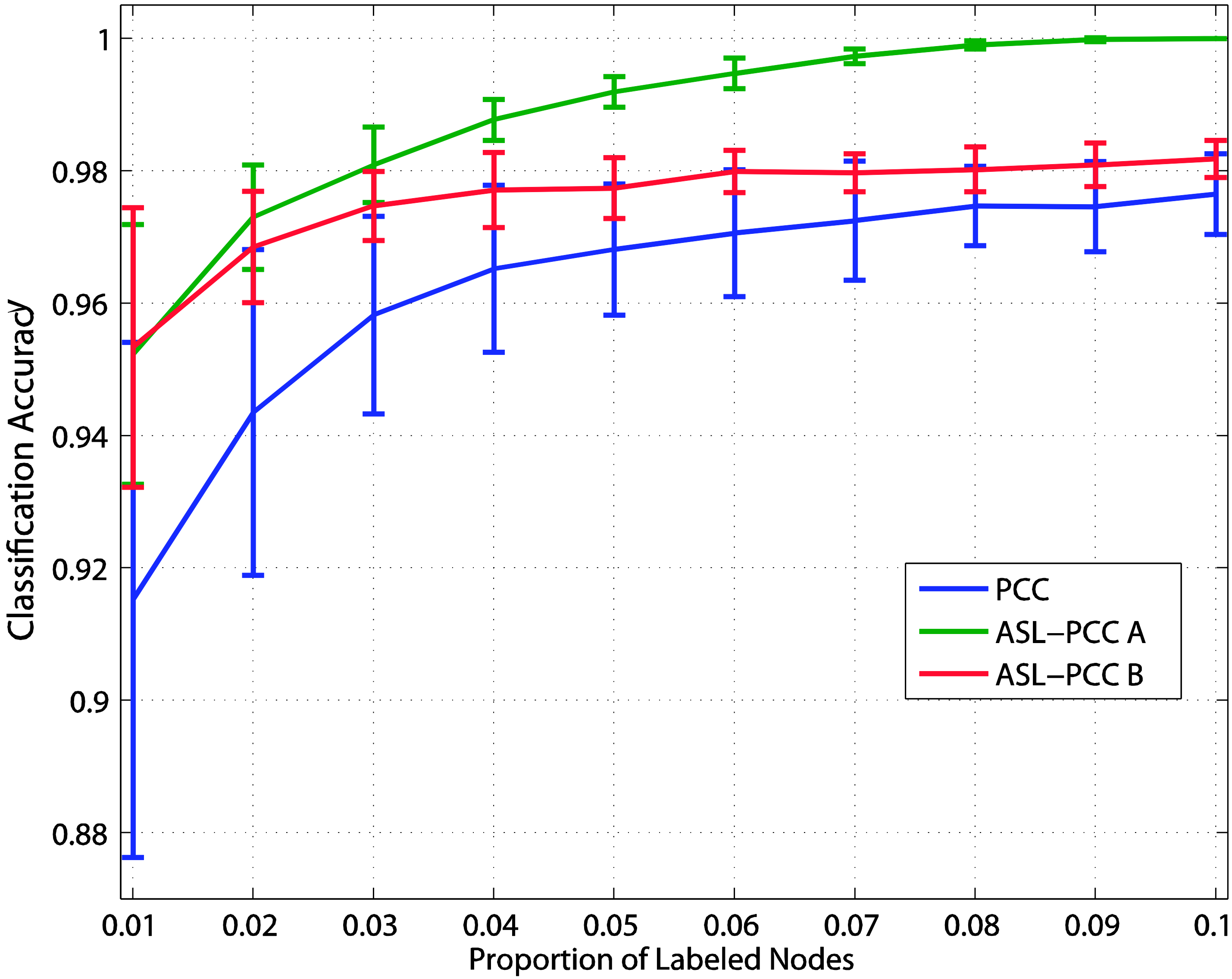
- Original PCC method
  - 1% to 10% labeled nodes are randomly chosen.
- ASL-PCC A and ASL-PCC B
  - Only 1 labeled node from each class is randomly chosen.
  - New query each time the system stabilizes.
    - Until it reaches 1% to 10% of labeled nodes.



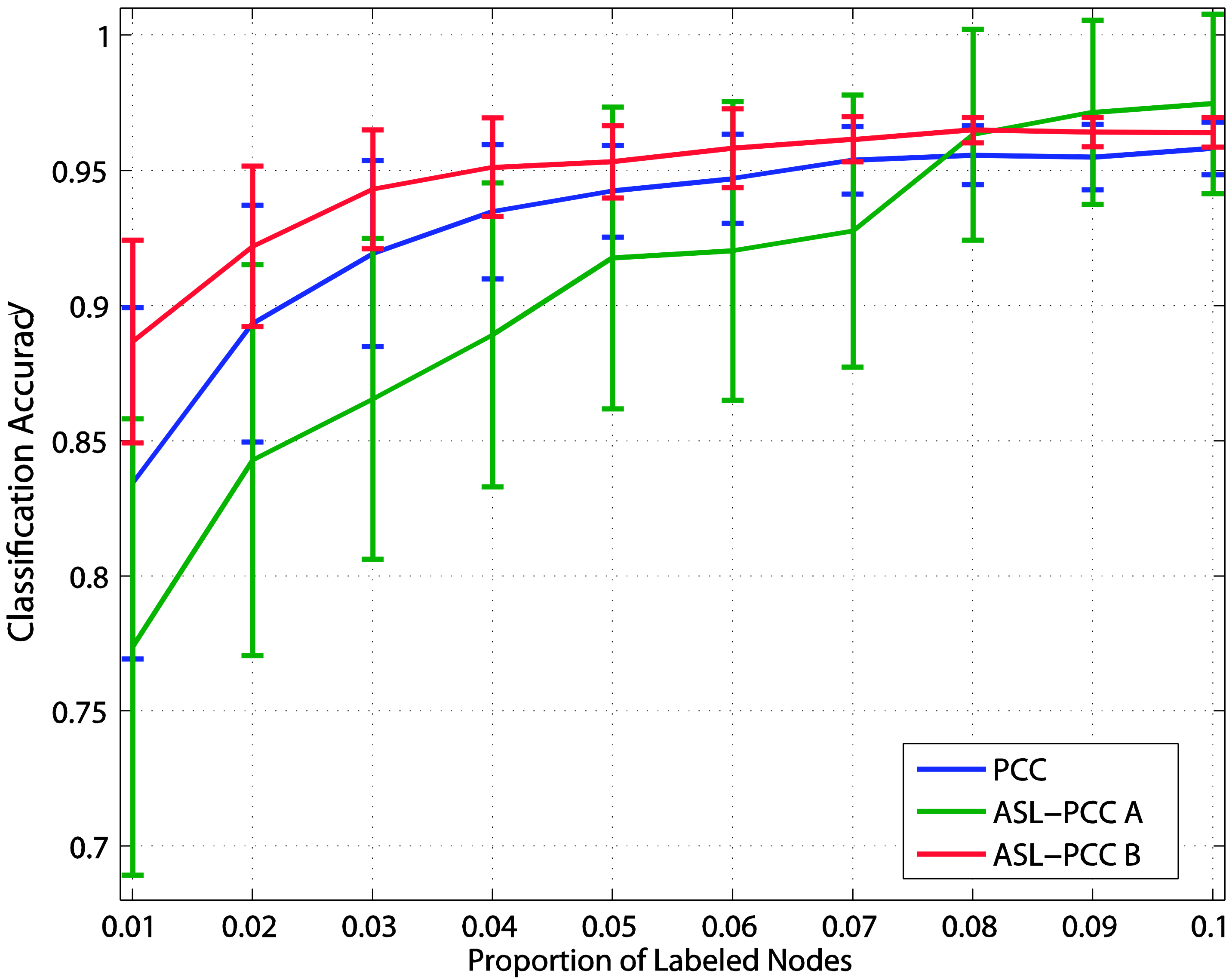
Correct classification rate comparison when the methods are applied to the Iris data set with different amounts of labeled nodes.



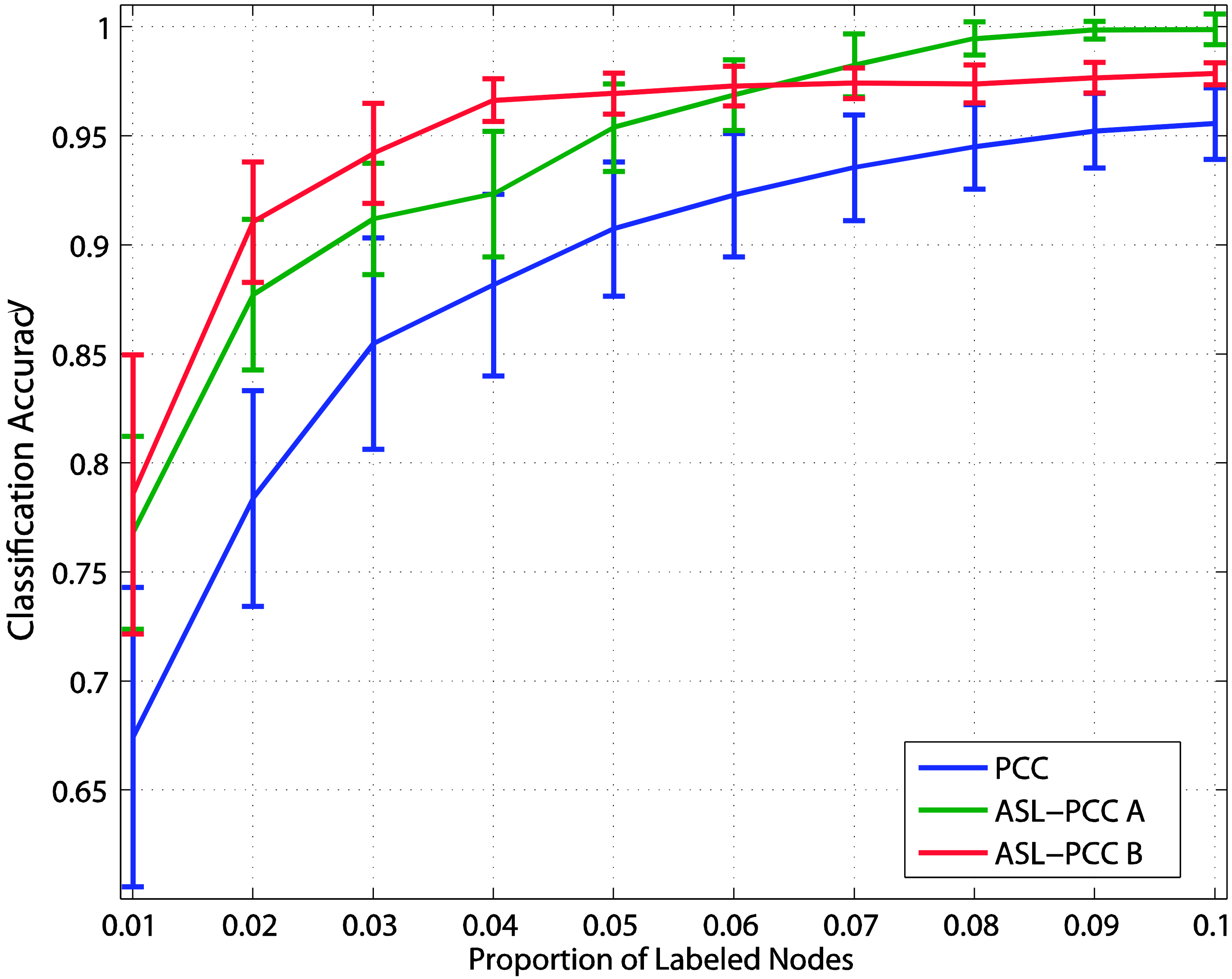
Correct classification rate comparison when the methods are applied to the Wine data set with different amounts of labeled nodes.



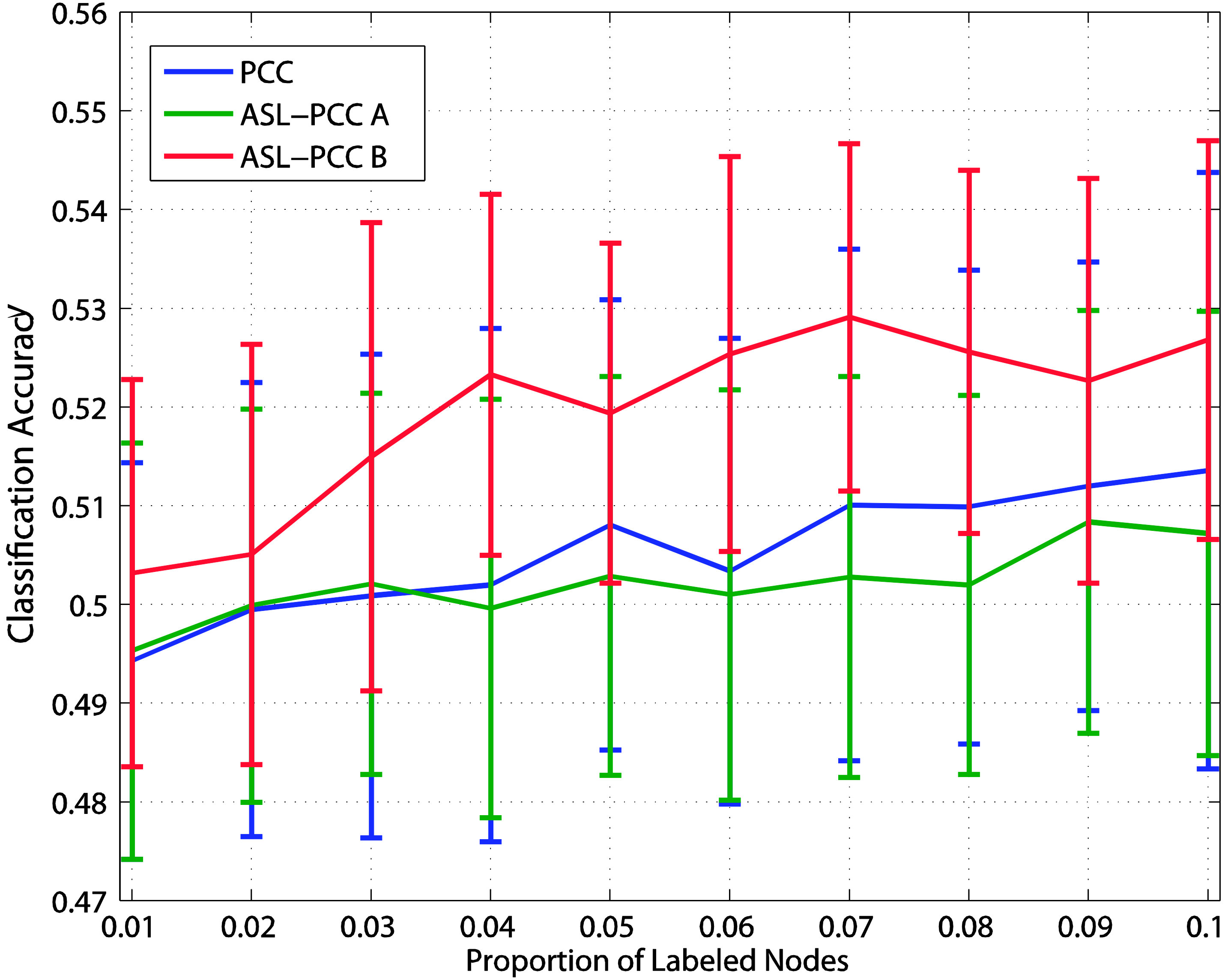
Correct classification rate comparison when the methods are applied to the Digit1 data set with different amounts of labeled nodes.



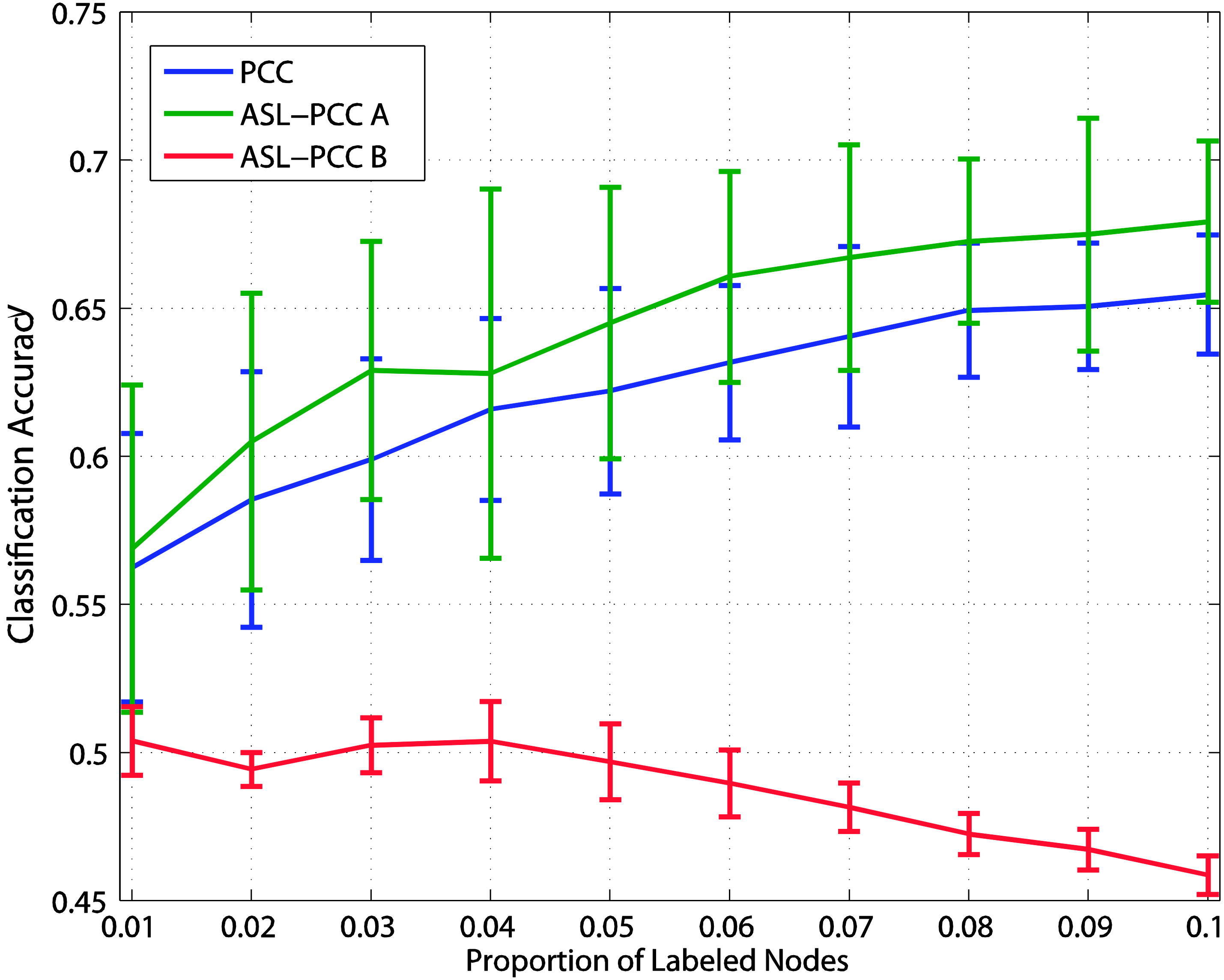
Correct classification rate comparison when the methods are applied to the USPS data set with different amounts of labeled nodes.



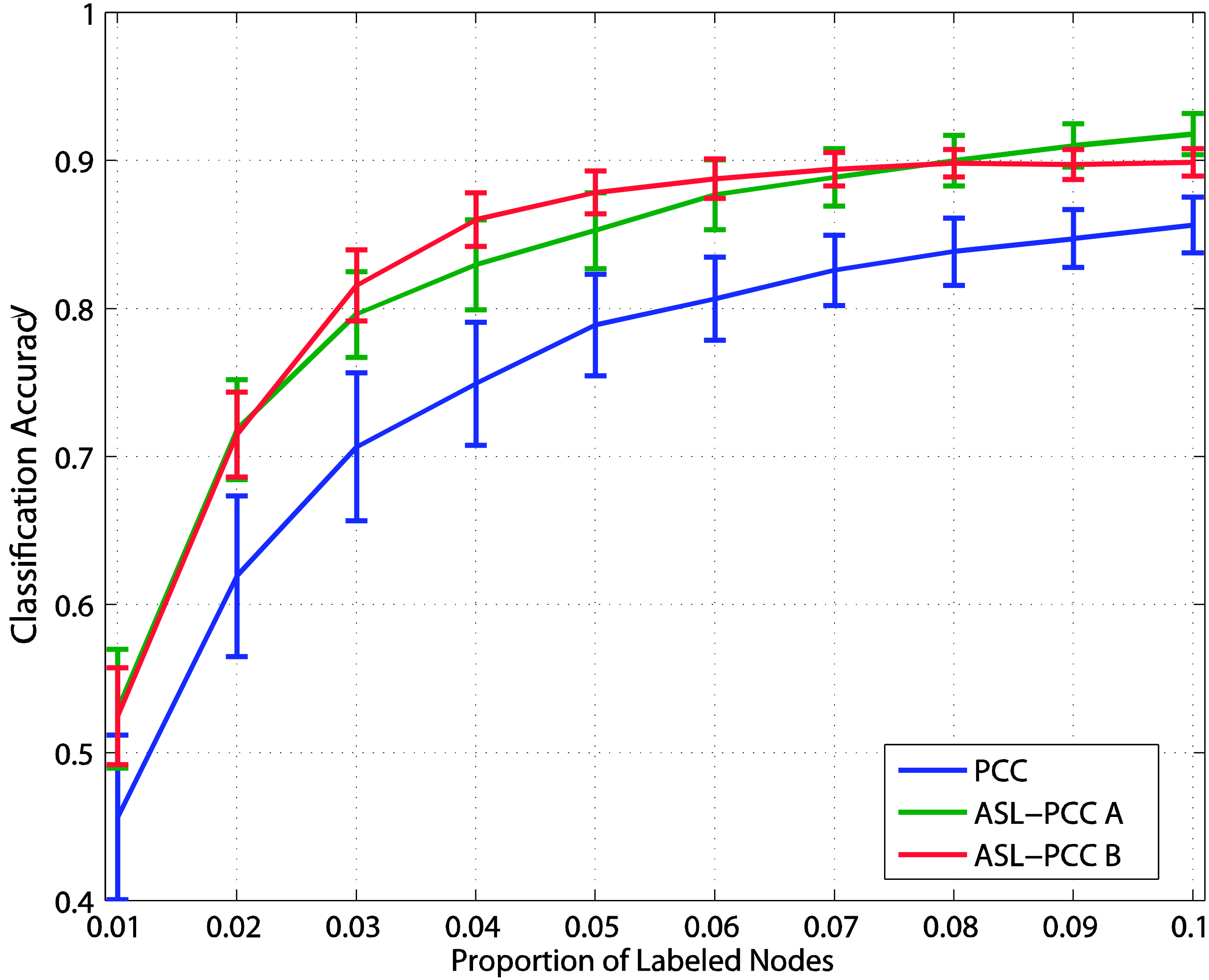
Correct classification rate comparison when the methods are applied to the $COIL_2$ data set with different amounts of labeled nodes.



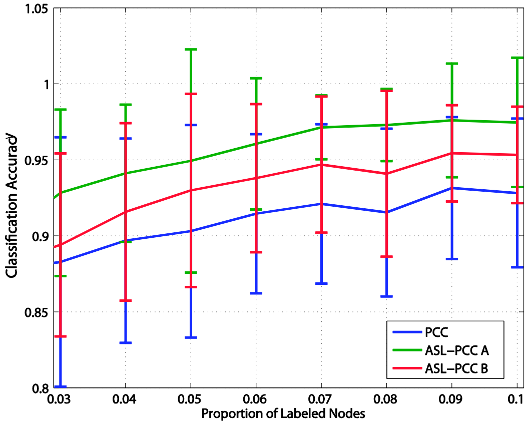
Correct classification rate comparison when the methods are applied to the BCI data set with different amounts of labeled nodes.

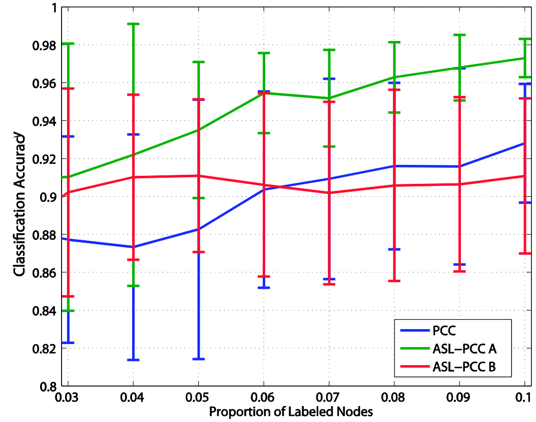


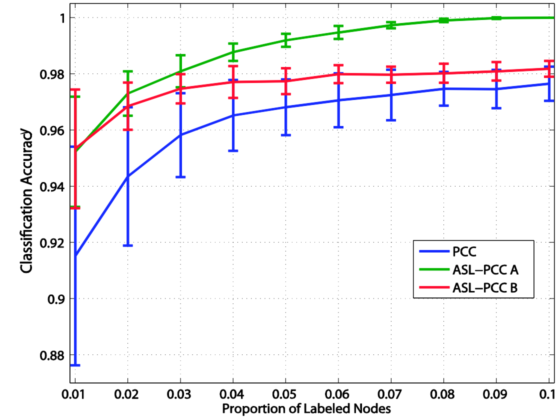
Correct classification rate comparison when the methods are applied to the g241c data set with different amounts of labeled nodes.

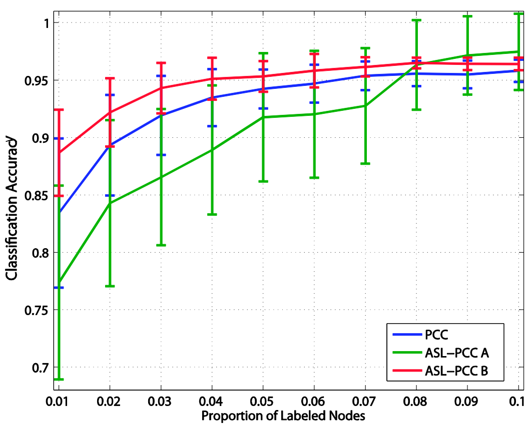


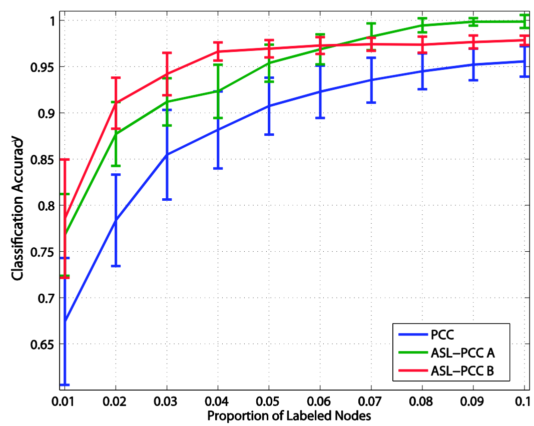
Correct classification rate comparison when the methods are applied to the COIL data set with different amounts of labeled nodes.

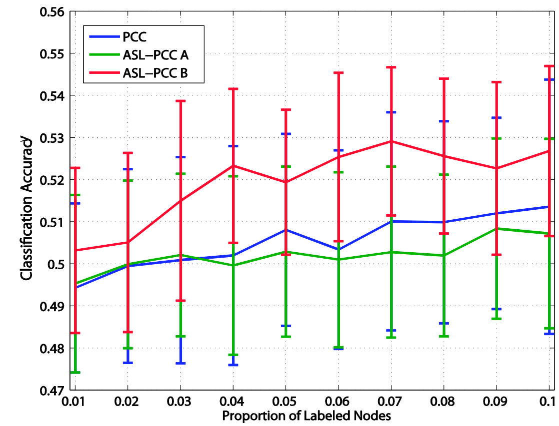

Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B

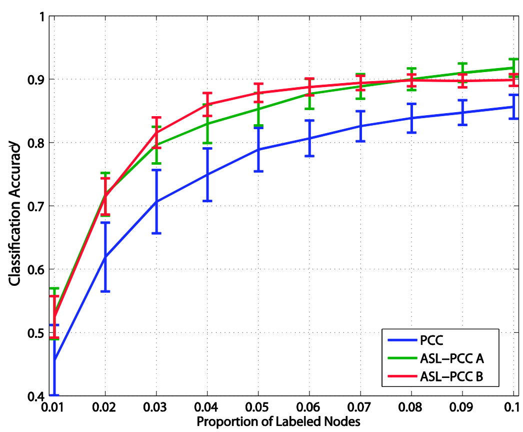

Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B

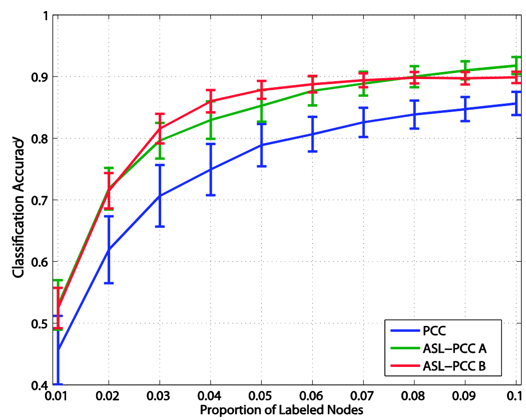

Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B


Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B


Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B


Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B


Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B


Classification Accuracy
Proportion of Labeled Nodes
PCC
ASL–PCC A
ASL–PCC B

Conclusions

- Semi-supervised learning and active learning features combined into a single approach
- Inspired on the collective behavior of social animals
  - Protect their territories against intruding groups.
- No Retraining
  - New particles are created on the fly as unlabeled nodes become labeled nodes.
  - The algorithm naturally adapts itself to new situations.
  - Only nodes affected by the new particles are updated
    - Equilibrium state is quickly reached again
    - Saves execution time.

Conclusions

- Better classification accuracy than the only semi-supervised learning counterpart when the same amount of labeled data is used.
  - **ASL-PCC A** is indicated when:
    - Classes are well separated.
    - Frontiers do not have many outliers.
  - **ASL-PCC B** is indicated when:
    - Frontiers are not well defined.
    - There are overlapped regions.
    - There are many outliers.

1st BRICS Countries Congress (BRICS-CCI) and 11th Brazilian Congress (CBIC) on Computational Intelligence

Combined Active and Semi-Supervised Learning using Particle Walking Temporal Dynamics

Fabricio Breve fabricio@rc.unesp.br

Department of Statistics, Applied Mathematics and Computation (DEMAC), Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (UNESP), Rio Claro, SP, Brazil