Steiner trees in the stochastic mean-field model of distance

Angus Davidson
University of Bristol

Joint work with Ayalvadi Ganesh
and Balint Toth

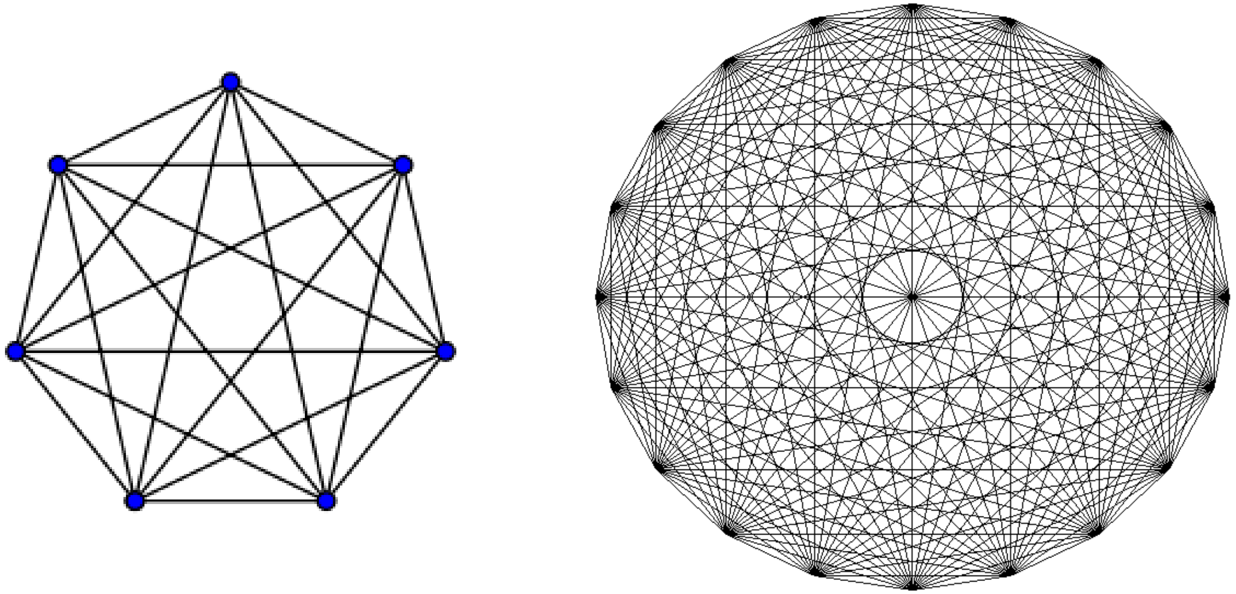
Climbing Redwoods

Angus Davidson
University of Bristol

Joint work with Ayalvadi Ganesh
and Balint Toth

Random edge-weighted graphs

- Complete graph on n nodes, $G = K_n$
- Random edge weights, iid $Exp(1)$



Distance between nodes: Epidemic

- First passage percolation
- Single node infected initially
- Edge weight = time for infection to cross that edge
- Length of shortest path between nodes *u* and *v* is the same as the time for infection started at *u* to reach *v* (or vice versa)

Analysis of first-passage percolation

- T_k : first time that k nodes are infected
- Number of edges between infected and uninfected nodes : $k(n-k)$
- Time to infect one more node is minimum of $k(n-k)$ independent $Exp(1)$ r.v.s.

$$E[T_{k+1} - T_k] = \frac{1}{k(n-k)} = \frac{1}{n}\left(\frac{1}{k} + \frac{1}{n-k}\right)$$

Analysis (cont.)

- Time to infect all nodes is

$$T_n = (T_n - T_{n-1}) + \cdots + (T_2 - T_1) + T_1$$

- So $E[T_n] \sim 2\log(n)/n$
- Most nodes infected $\sim \log(n)/n$
- Diameter of graph $\sim 3\log(n)/n$

Steiner tree problem

- Fix k points on our graph G
- Find the minimum weight tree connecting the points
- For k typical nodes call this weight W_k
- Study asymptotics of this random variable, for k fixed and n tending to infinity

Previous results

- Bollobas, Gamarnik, Riordan and Sudakov (2004):

$$W_k \sim (k-1)\frac{\log(n)-\log(k)}{n}$$

- Here, the k nodes are chosen at random. Equivalently, the nodes are fixed first, and the edge weights assigned afterwards.

Previous results

- Bollobas, Gamarnik, Riordan and Sudakov (2004):

$$W_k \sim (k-1)\frac{\log(n)}{n}$$

- Here, the k nodes are chosen at random. Equivalently, the nodes are fixed first, and the edge weights assigned afterwards.

Previous results

- What if we first assign edge weights, then choose the k nodes that maximise the weight of the Steiner tree?
- Call this random variable M_k
- Janson (1999):

$$W_2 \sim \frac{\log(n)}{n}, \qquad M_2 \sim 3\frac{\log(n)}{n}$$

More precisely

$$\frac{W_2}{\log n/n} \xrightarrow{p} 1 \qquad as\ n \to \infty$$

$$\frac{M_2}{\log n/n} \xrightarrow{p} 3 \qquad as\ n \to \infty$$

$$\frac{W_k}{\log n/n} \xrightarrow{p} (k-1) \quad as\ n \to \infty$$

Distribution of Typical Distance

$$\frac{W_2}{\log n/n} \xrightarrow{p} 1 \ \ as \ n \to \infty$$

$$nW_2 - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_3 \ \ as \ n \to \infty$$

Related work

- Random edge-weighted model very popular, has been used to study many combinatorial optimisation problems:

- Minimum spanning tree (Frieze, 1985):

$$W_n = M_n \sim \zeta(3) = \sum_{j=1}^{\infty} \frac{1}{j^3}$$

Related work

- Travelling salesman tour (Frieze, 2004):

$$\zeta(3) \leq W_{TSP} \leq 6$$

- Shortest path tree (van der Hofstad, Hooghiemstraa, van Mieghem, 2006):

$$W_{SPT} \sim \zeta(2) = \sum_{j=1}^{\infty} \frac{1}{j^2}$$

- Distribution of Diameter (Bhamidi, van der Hofstad, 2013)

Our results (1)

- The weight of the max-weight k-Steiner tree among all choices of k nodes is

$$M_k \sim (2k-1)\frac{\log(n)}{n}$$

Our results (1)

- Theorem (D., A. Ganesh)

$$\frac{M_k}{\log n/n} \xrightarrow{p} (2k-1) \;\; as \; n \to \infty$$

Intuition

- Typical distance between most pairs of nodes is $\log(n)/n$
- Some nodes are remote – at distance $\log(n)/n$ from their nearest neighbour
- But have ‘typical’ neighbours at this distance
- The typical neighbours are joined by a k-Steiner tree of weight $(k-1)\log(n)/n$
- Intuition can be turned into a lower bound

Some loose upper bounds

- Janson's result on graph diameter implies that
$$M_k \leq 3(k-1)\frac{\log(n)}{n}$$
- Consider infection started at a typical node: reaches all nodes by time $2\log(n)/n$
- Use resulting paths to connect given k nodes: yields
$$M_k \leq 2k\,\frac{\log(n)}{n}$$

Sketch of upper bound proof

- Want to show that, for any k nodes, the weight of the k-Steiner tree connecting them is bounded by $(2k-1)\frac{\log(n)}{n}$
- Use Chernoff bound on RV dominating weight of a typical k-Steiner tree, apply union bound

Sketch of upper bound proof

- Want to show that, for any k nodes, the weight of the k-Steiner tree connecting them is bounded by $(2k-1)\frac{\log(n)}{n}$
- Use Chernoff bound on RV dominating weight of a typical k-Steiner tree, apply union bound

$$
\begin{aligned}
& \mathbb{P}(W_k \geq (2k-1+\varepsilon)\log n/n) \\
\leq\ & \mathbb{P}(X(S) \geq (2k-1+\varepsilon)\log n/n) \\
\leq\ & \mathbb{E}(e^{Xnt-(2k-1+\varepsilon)t\log n}) \\
=\ & O(n^{-k-\varepsilon})
\end{aligned}
$$

Sketch of upper bound proof

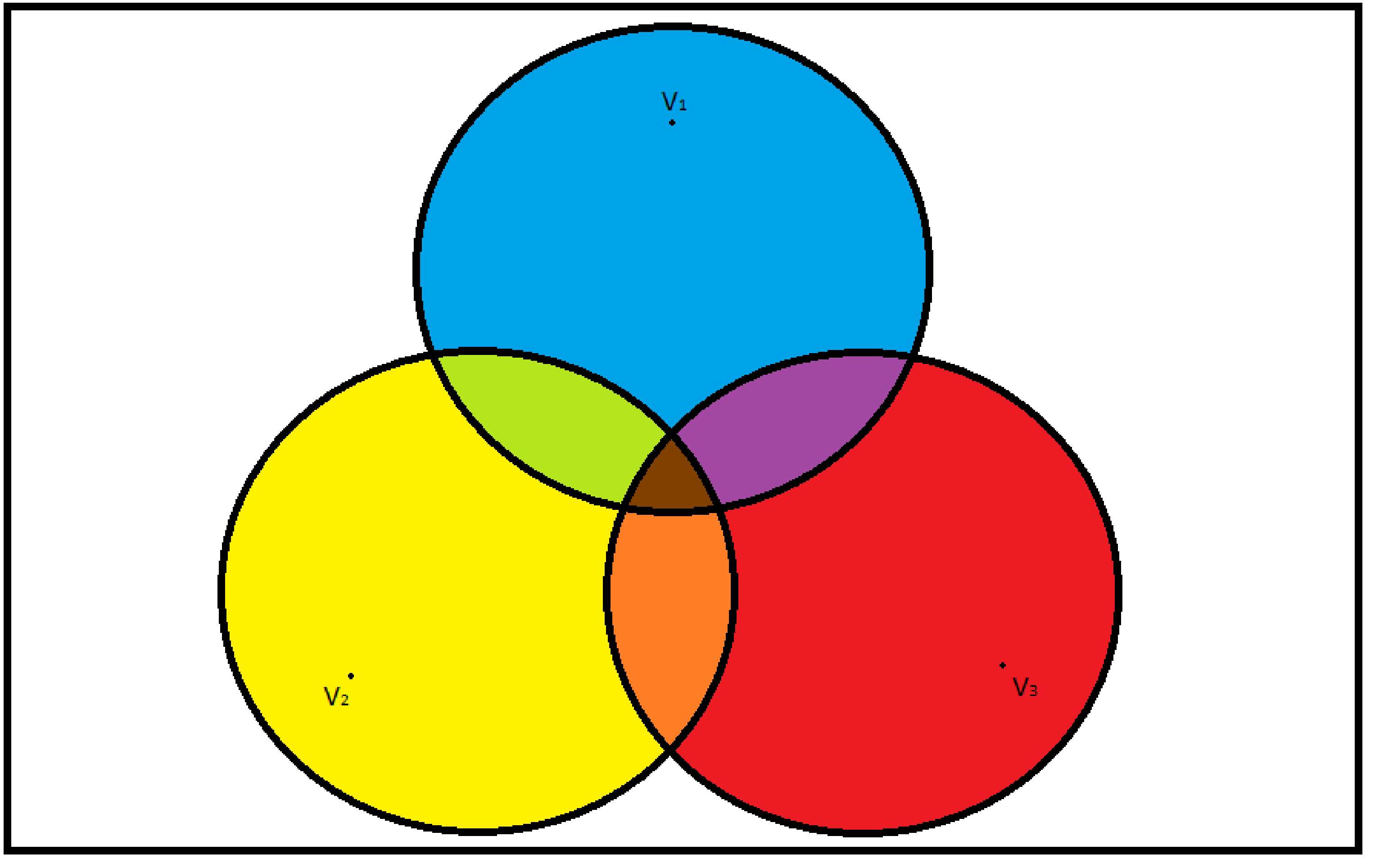
$$\begin{aligned}
& \mathbb{P}(W_k \geq (2k-1+\varepsilon)\log n/n) \\
\leq\ & \mathbb{P}(X(S) \geq (2k-1+\varepsilon)\log n/n) \\
\leq\ & \mathbb{E}(e^{Xnt-(2k-1+\varepsilon)t\log n}) \\
=\ & O(n^{-k-\varepsilon})
\end{aligned}$$

$$\begin{aligned}
& \mathbb{P}(M_k \geq (2k-1+\varepsilon)\log n/n) \\
\leq\ & \mathbb{P}\left(\bigcup_{|S|=k} X(S) \geq (2k-1+\varepsilon)\log n/n\right) \\
\leq\ & \binom{n}{k} O(n^{-k-\varepsilon}) \\
=\ & O(n^{-\varepsilon})
\end{aligned}$$

Upper bound: intuition

- Pick k nodes, and start infection simultaneously from each of them
- Grow infected sets until each has (a bit more than) $n^{(k-1)/k}$ nodes
- If subsets were independent, would all have a node in common

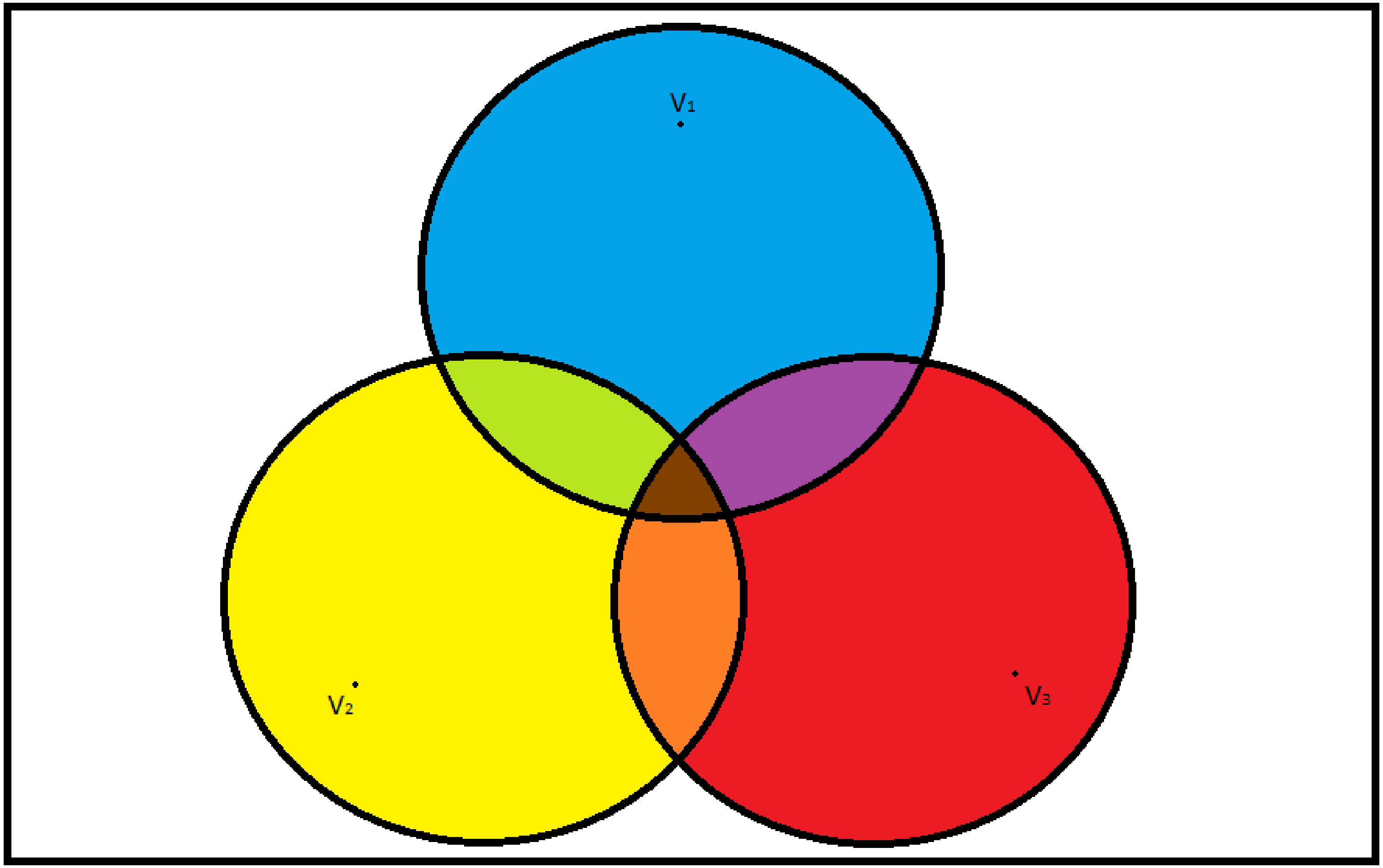
Upper bound: intuition



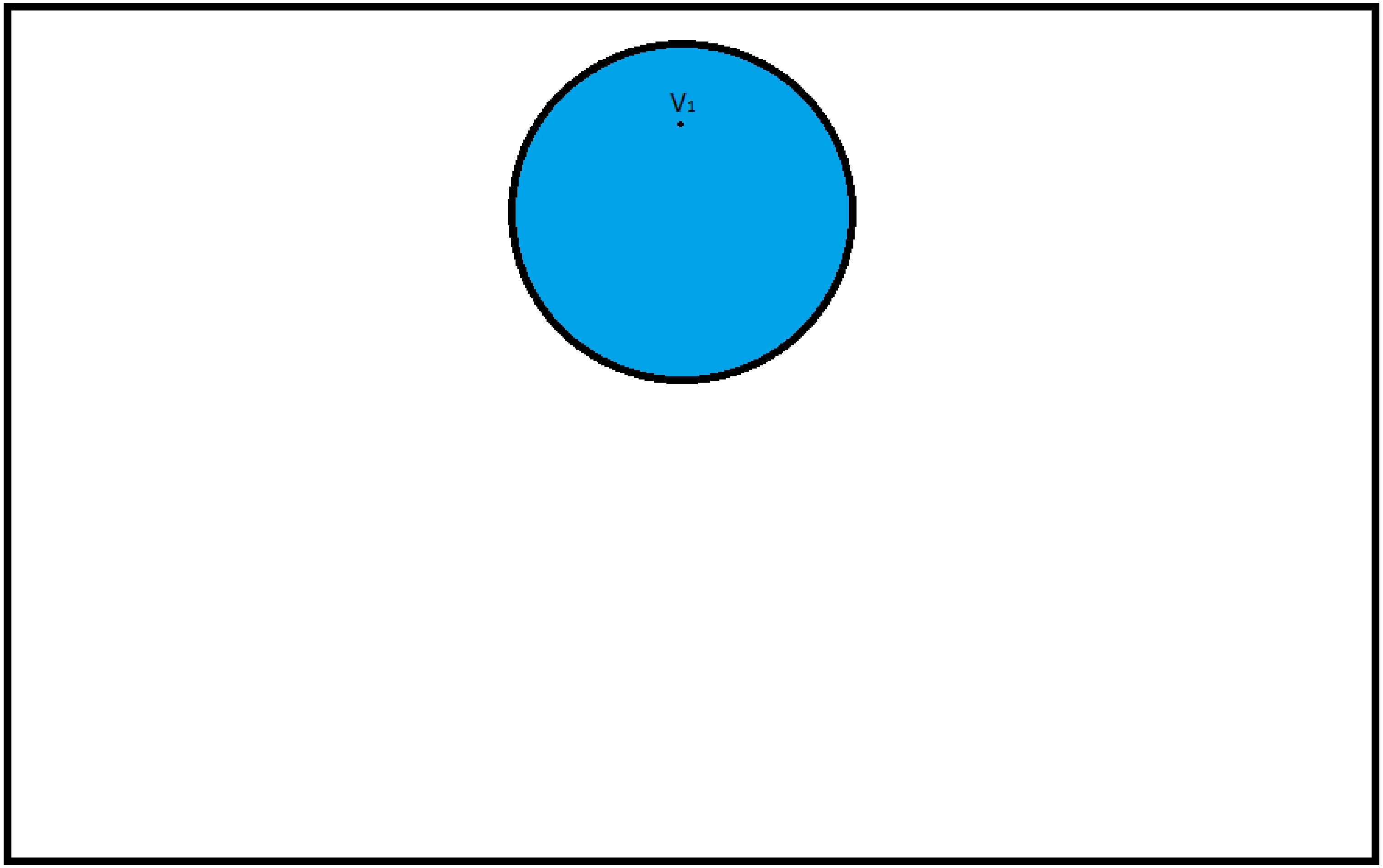
Upper bound: intuition

- Pick k nodes, and start infection simultaneously from each of them
- Grow infected sets until each has $Cn^{(k-1)/k}$ nodes
- If subsets were independent, would all have a node in common
- *But sets are not independent after first pair intersect*

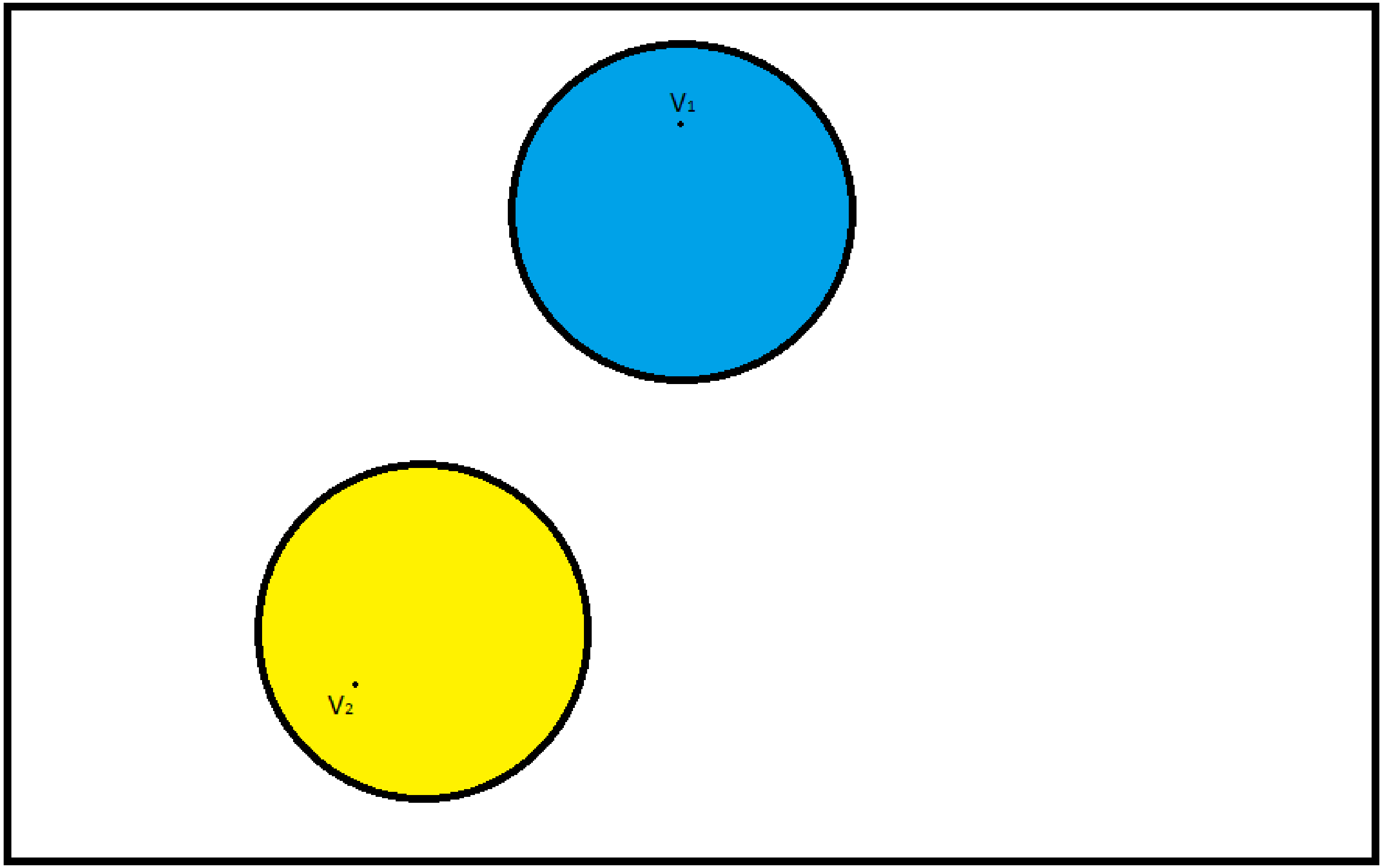
Upper bound: intuition



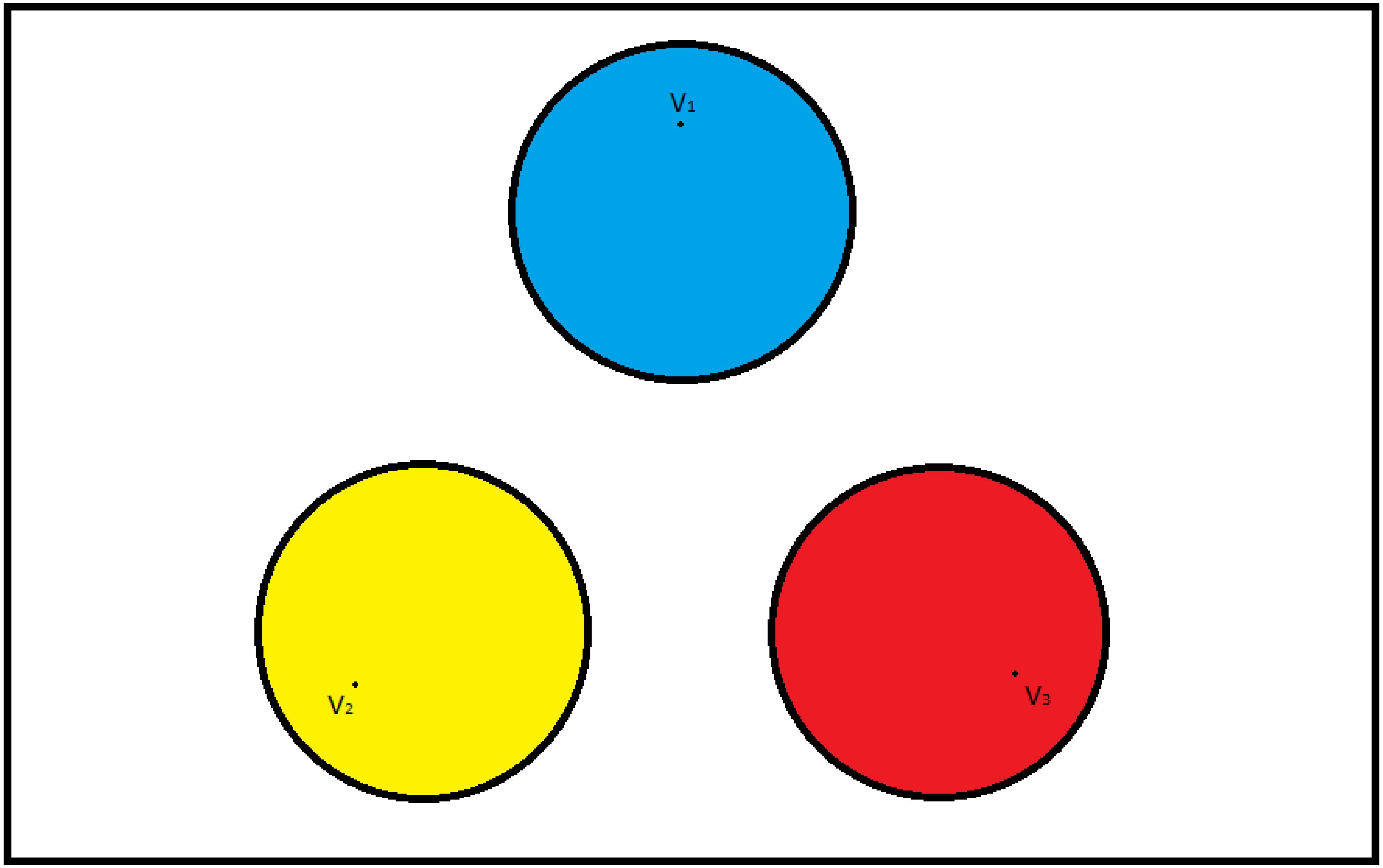
Upper bound: intuition



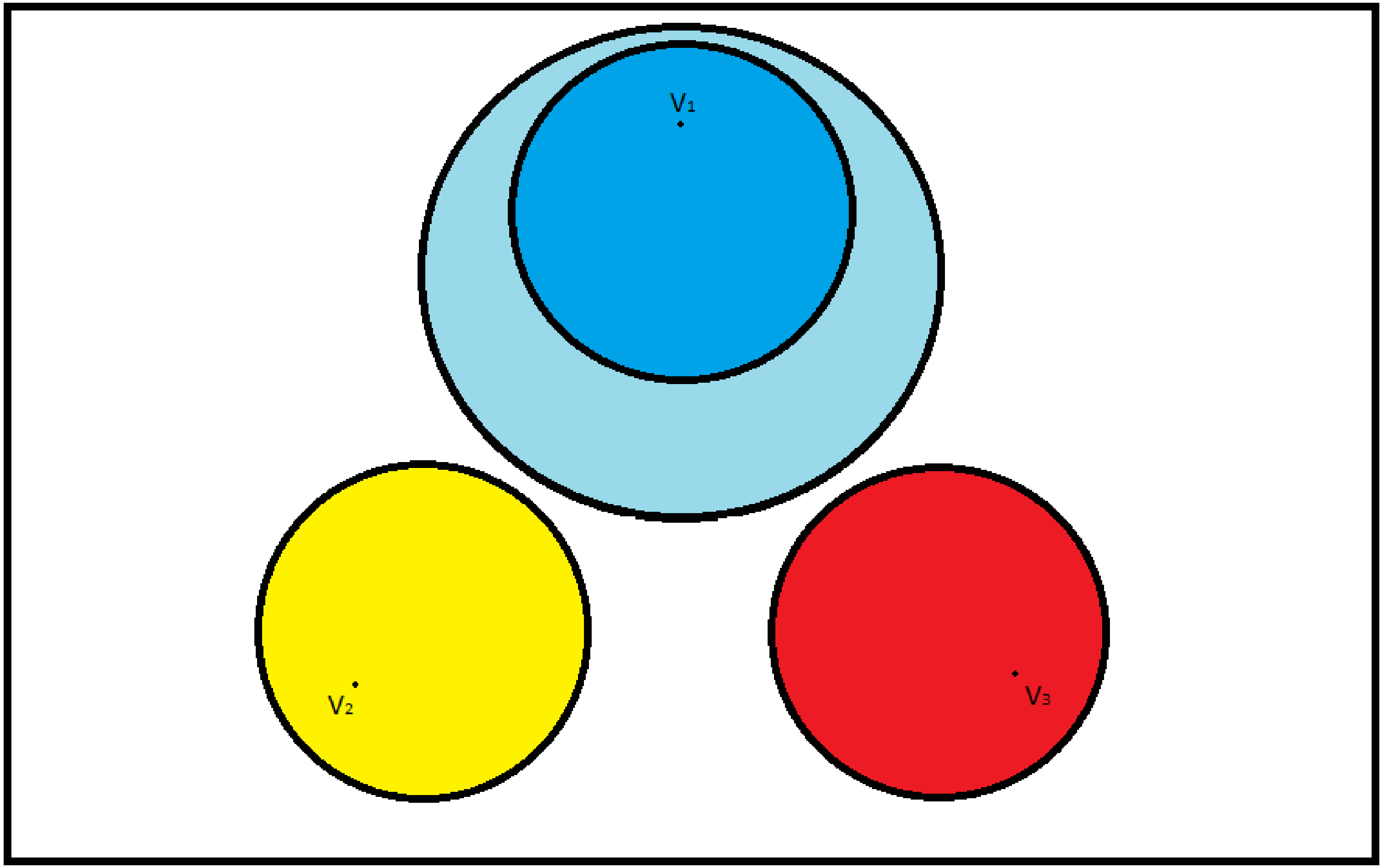
Upper bound: intuition



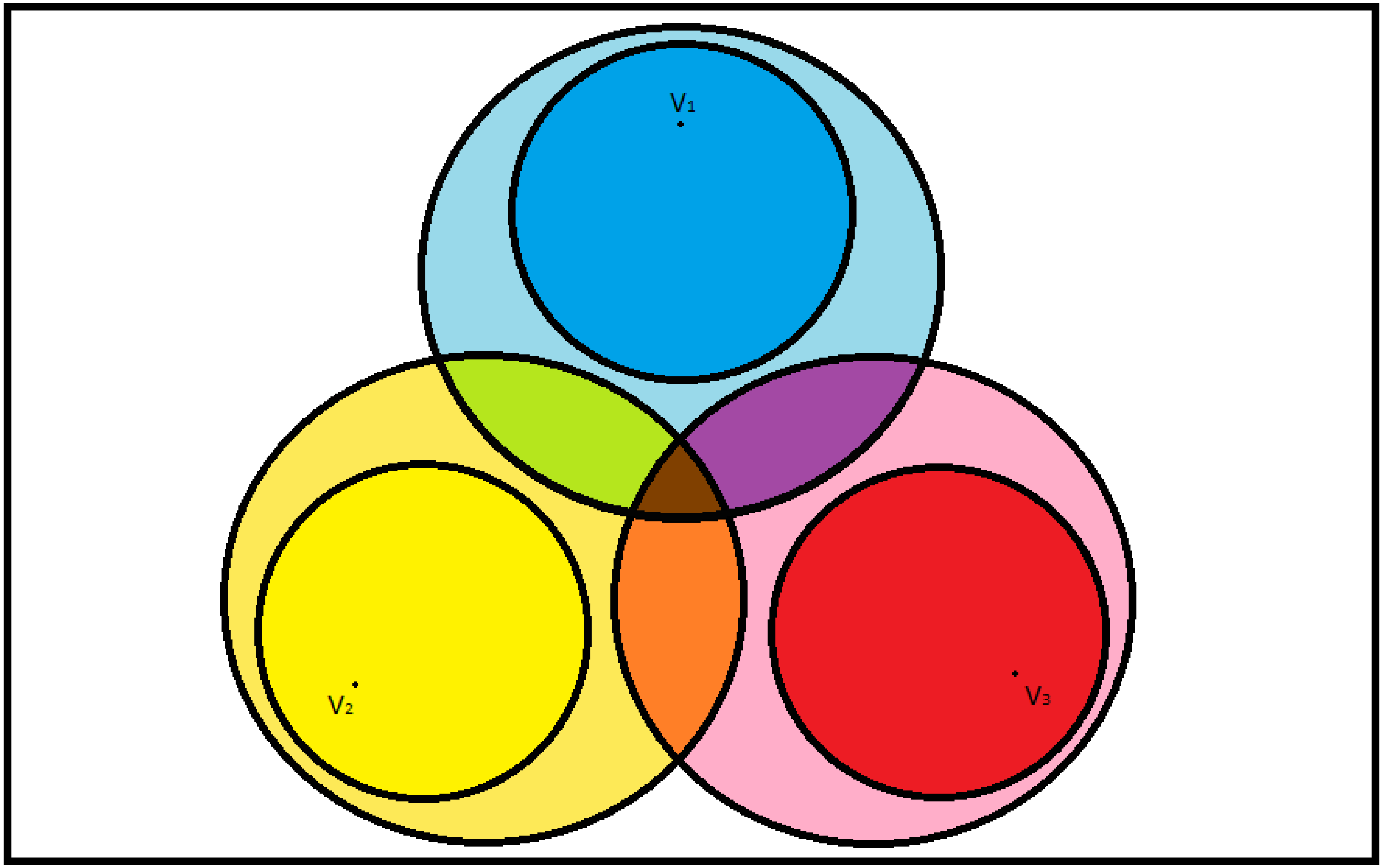
Upper bound: intuition



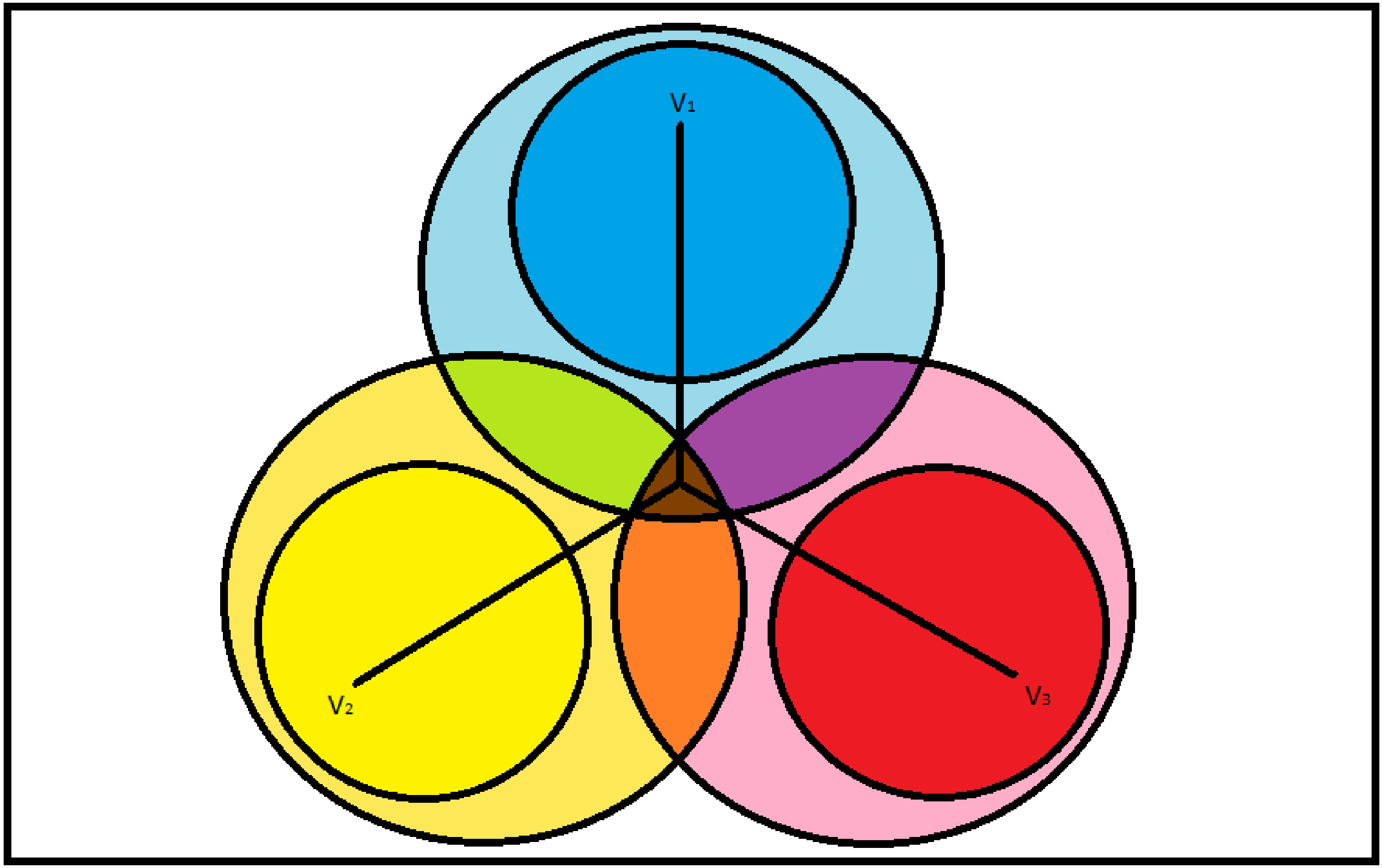
Upper bound: intuition



Upper bound: intuition



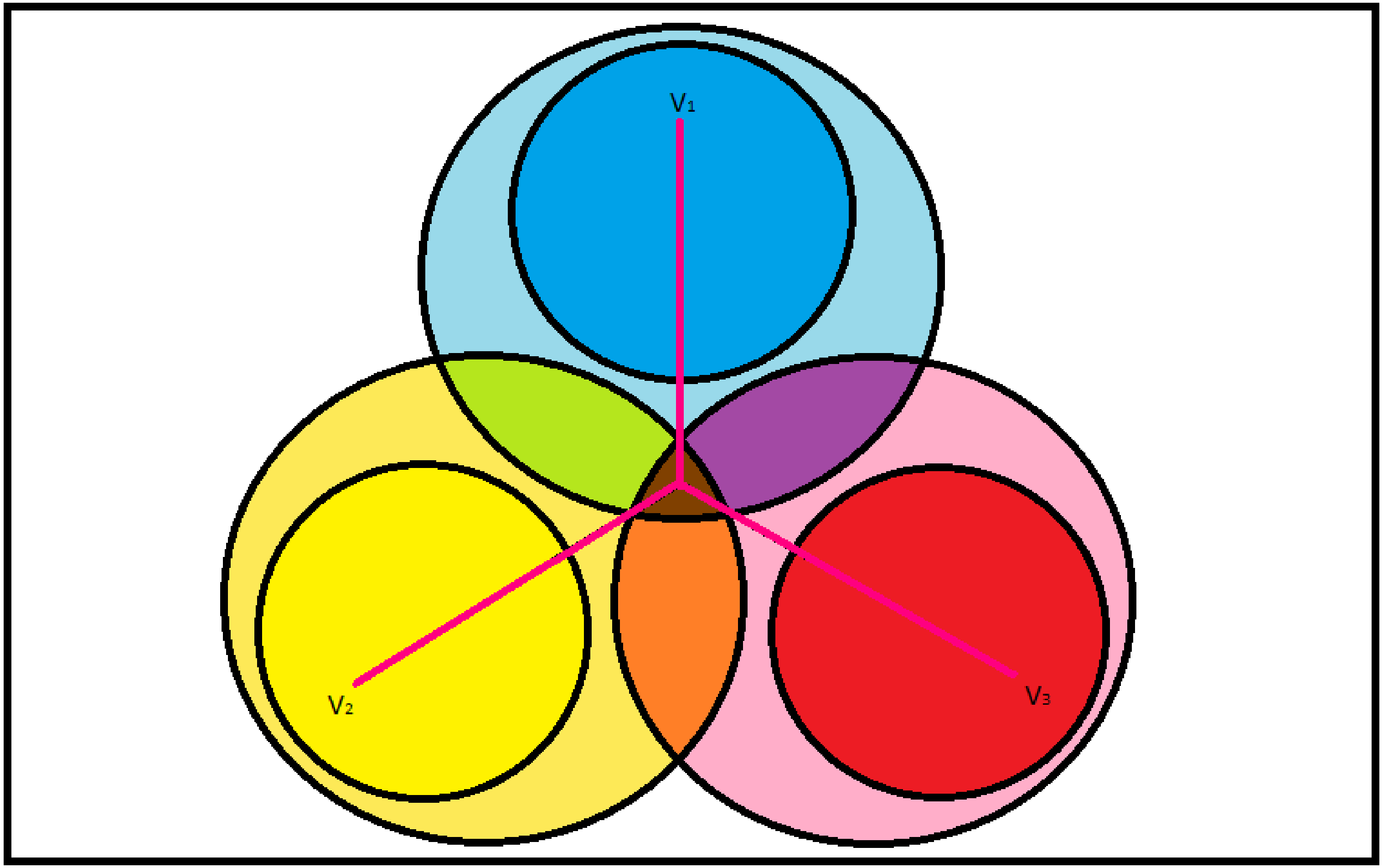
Upper bound: intuition



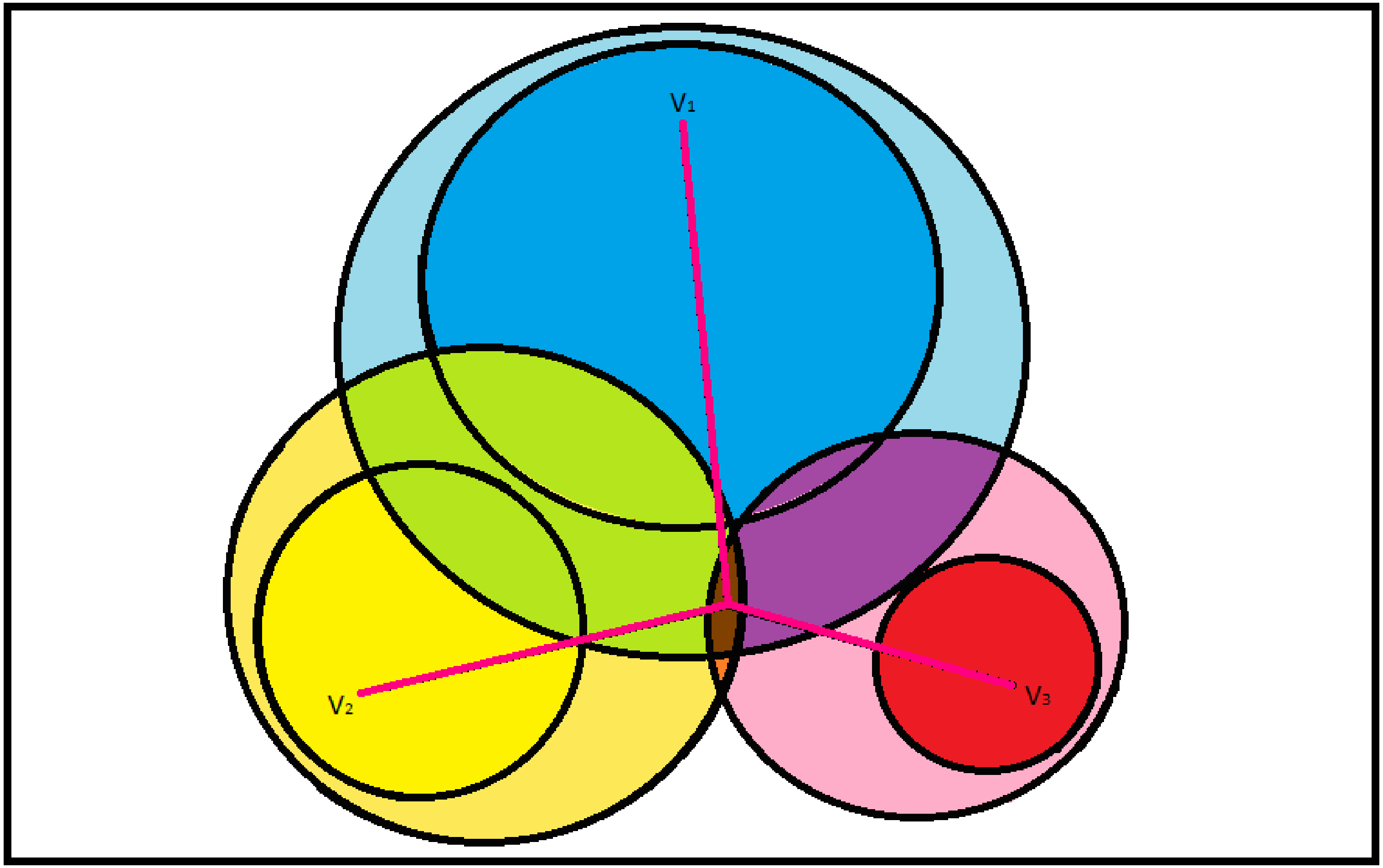
Asymptotic shapes

- Weight of k-Steiner tree is $(k-1)\log(n)/n$
- Same as sum of shortest path lengths from one of the k nodes to the other $k-1$
- Suggests tree is degenerate – has no internal nodes
- Growing infected sets from each node gives same total weight, but different shape
- Expected weight not sufficiently informative, need to look at fluctuations

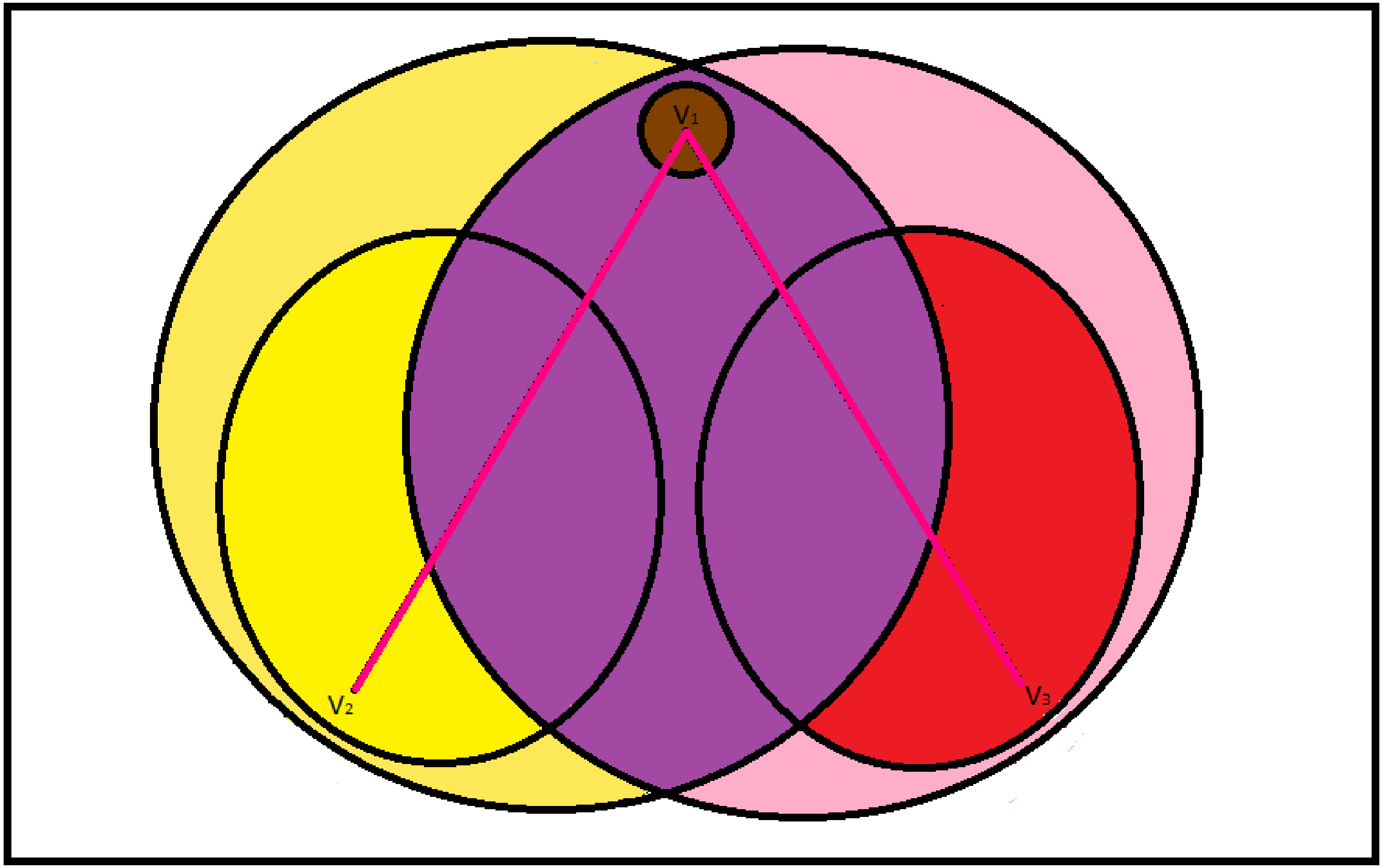
Asymptotic shapes



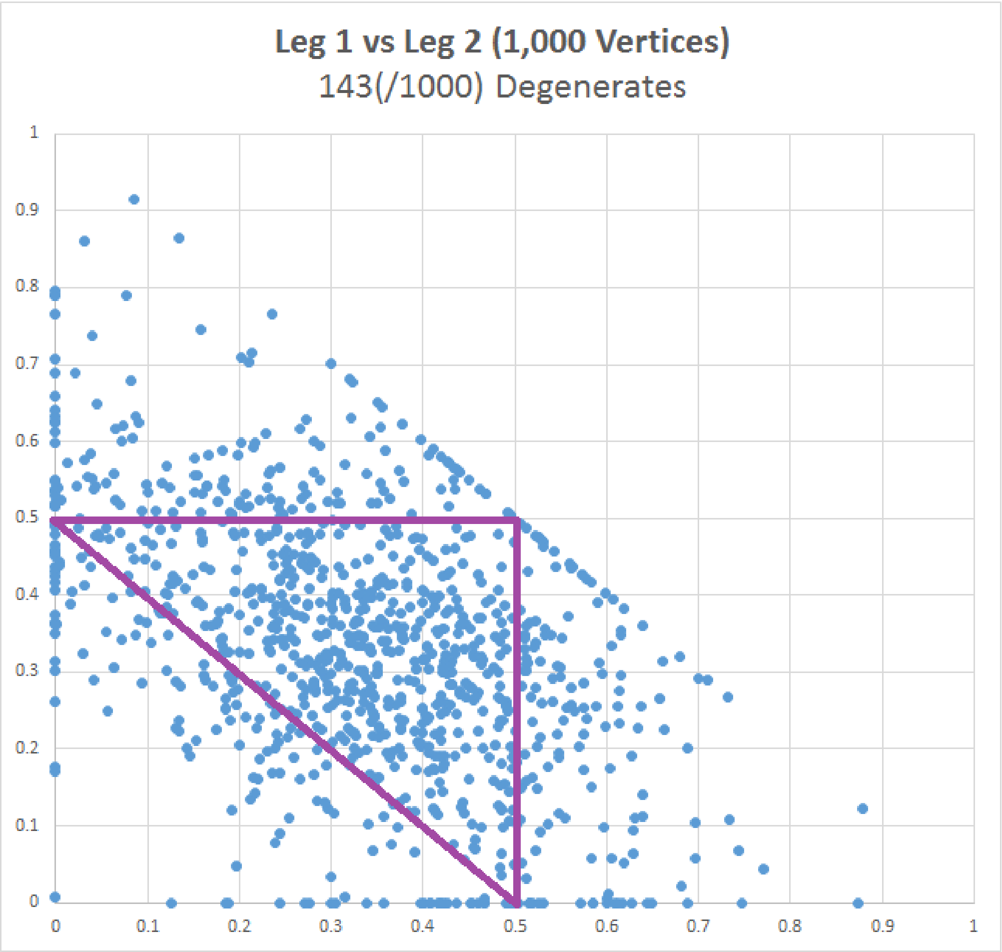
Asymptotic shapes



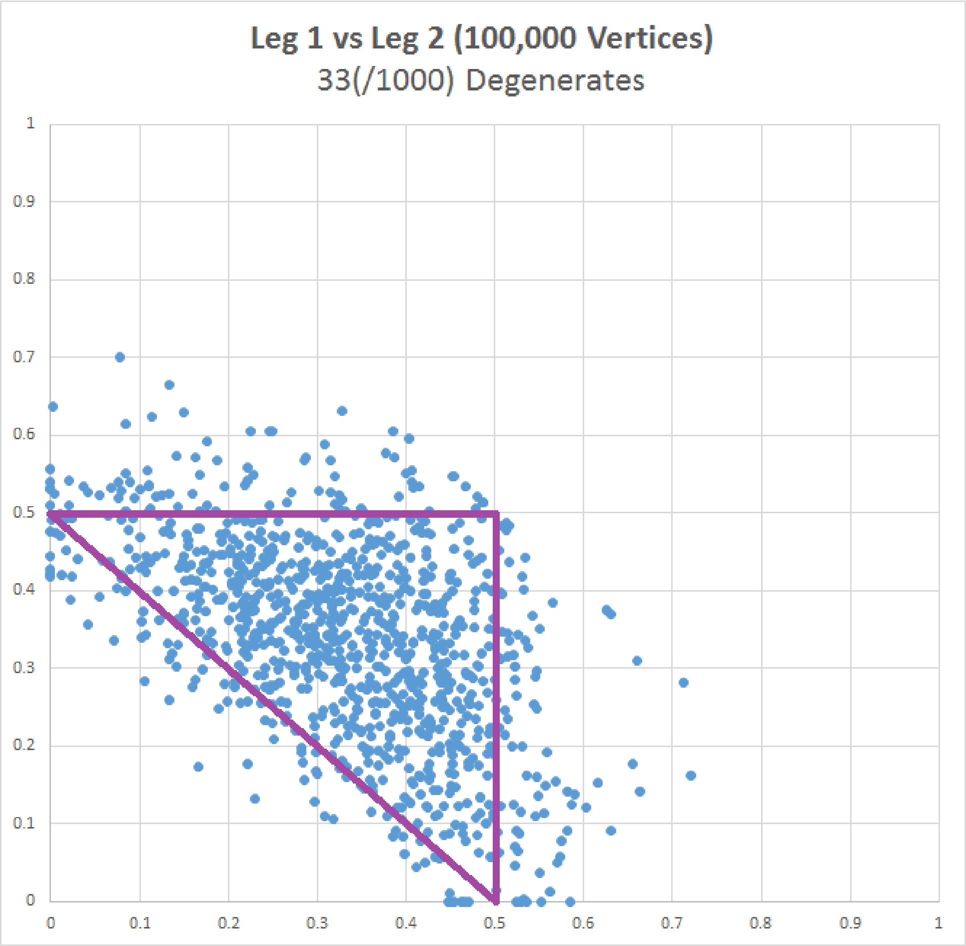
Asymptotic shapes



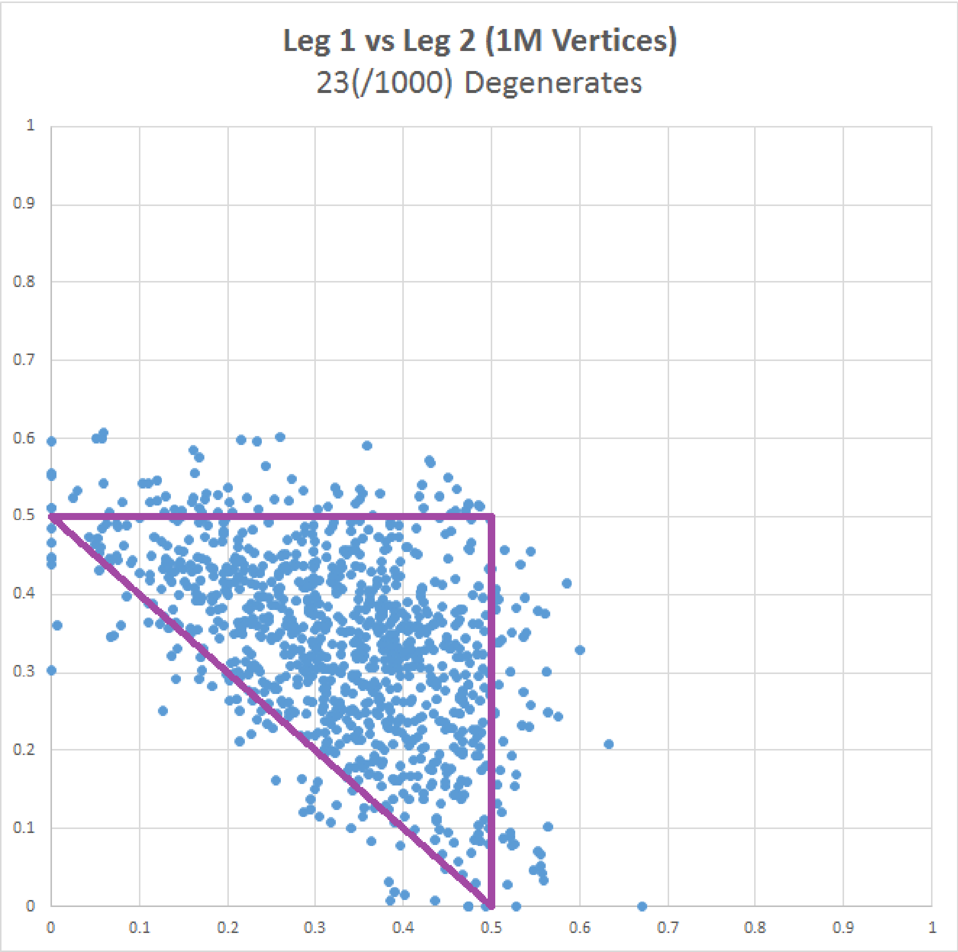
Simulations: 3-Steiner tree
2,250 Vertices



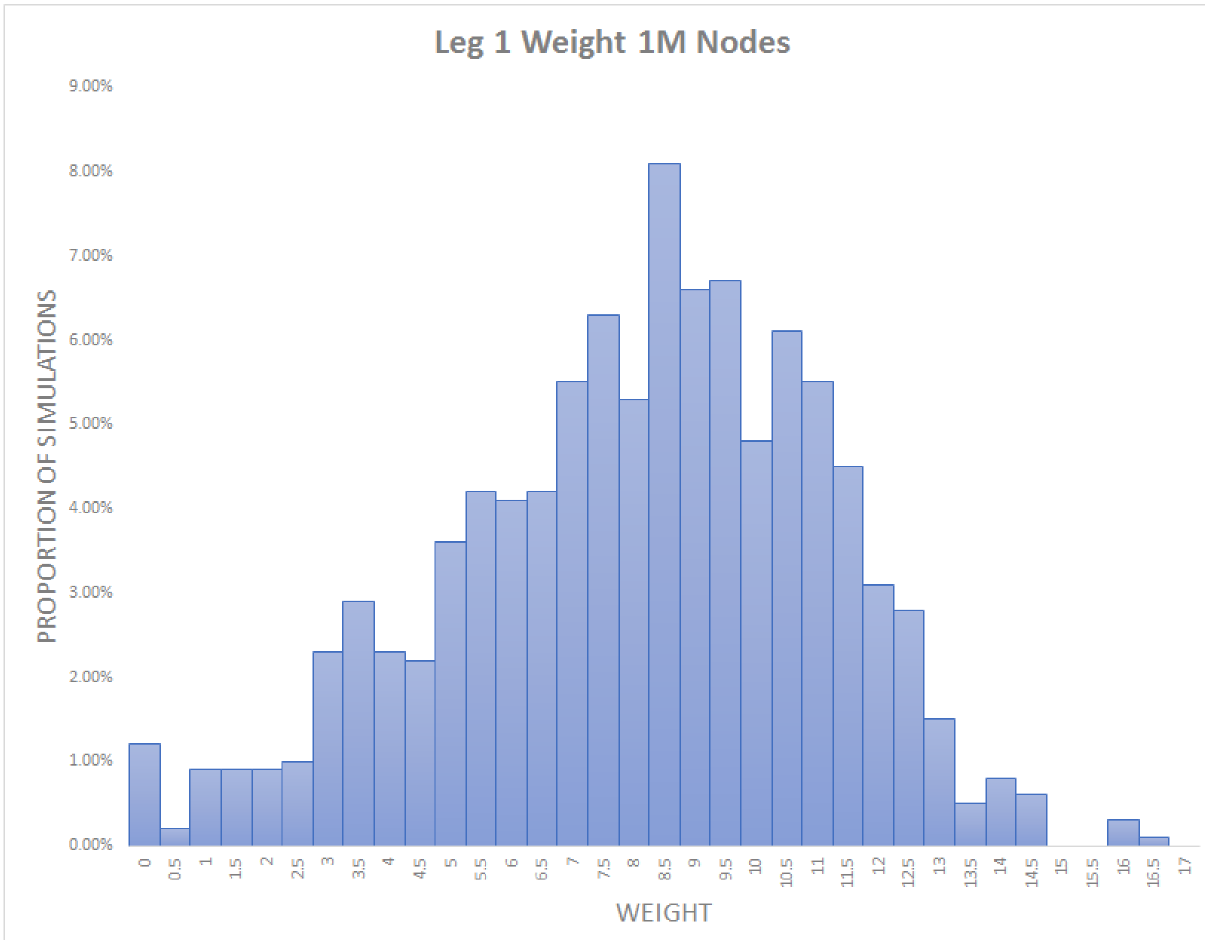
Simulations: 3-Steiner tree 100,000 Vertices



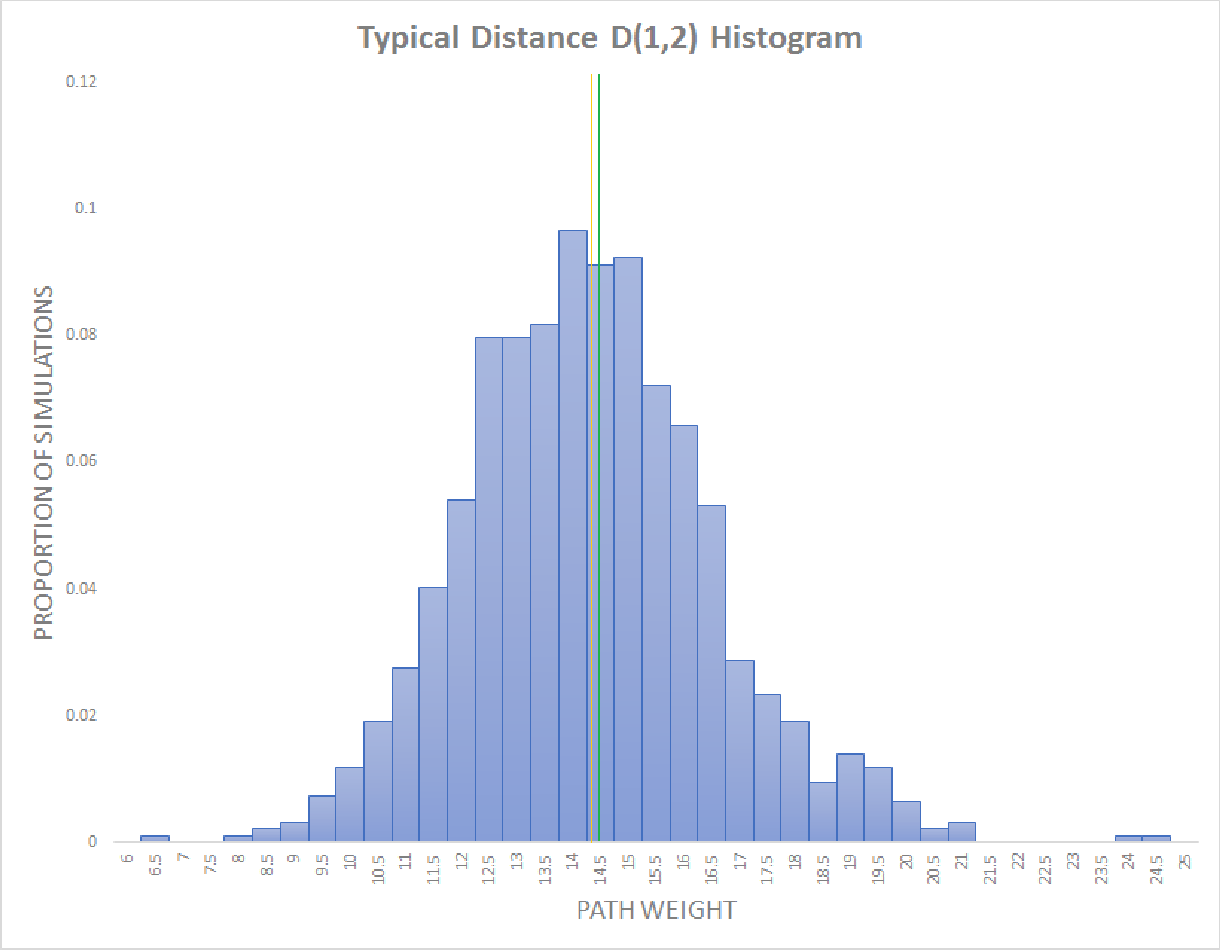
Simulations: 3-Steiner tree 1,000,000 Vertices



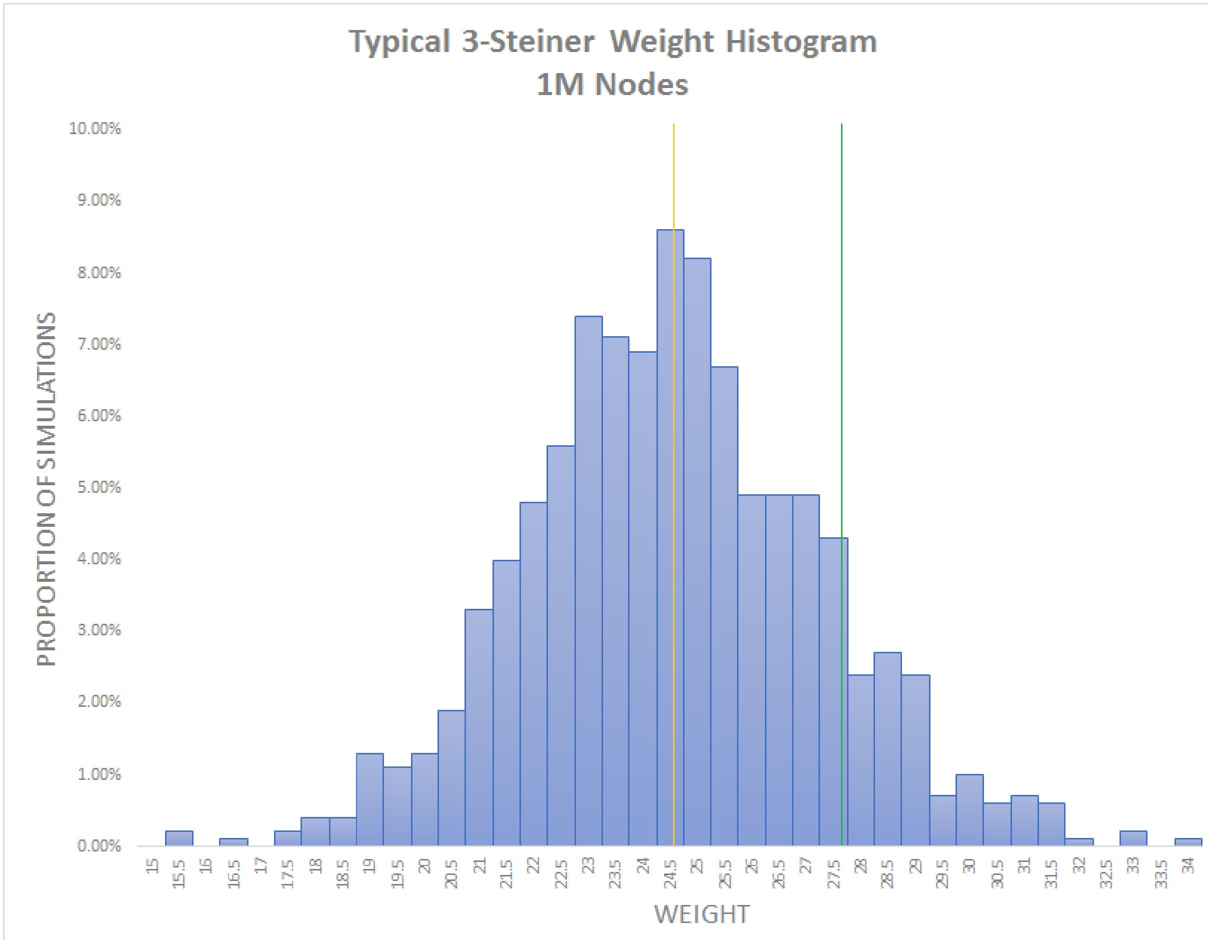
Leg 1 Distribution – 1M Vertices



Typical Distance Empirical Distribution cf. $\log(n) + \Lambda_1 + \Lambda_2 - \Lambda_{1,2}$



Typical 3-Steiner Tree Weight Empirical Distribution cf. 2log(n)



Average Tree Weight vs 2 log(n)

n	2log(n)	Average Tree Weight (1000 runs)	2log(n) - ATW
500	12.4	11.1	**1.35**
2250	15.4	13.7	**1.72**
100,000	23.0	20.5	**2.57**
1,000,000	27.6	24.5	**3.09**

Typical Steiner tree weight

- Conjecture:

$$nW_k - (k-1)\log n \xrightarrow{p} -\infty \ \ as\ n \to \infty$$

- Cf.

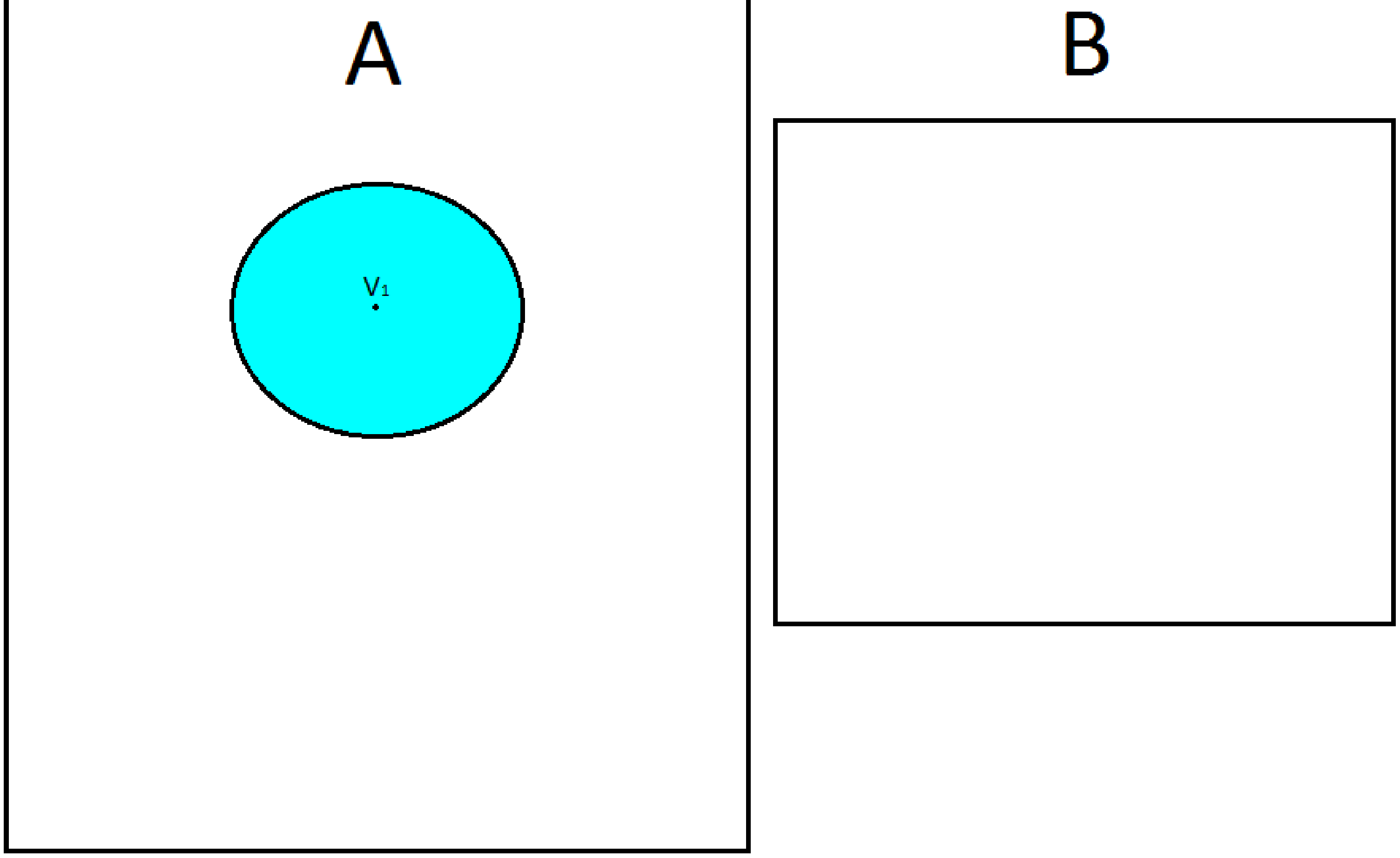
$$nW_2 - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_3 \ \ as\ n \to \infty$$

Our results (2)

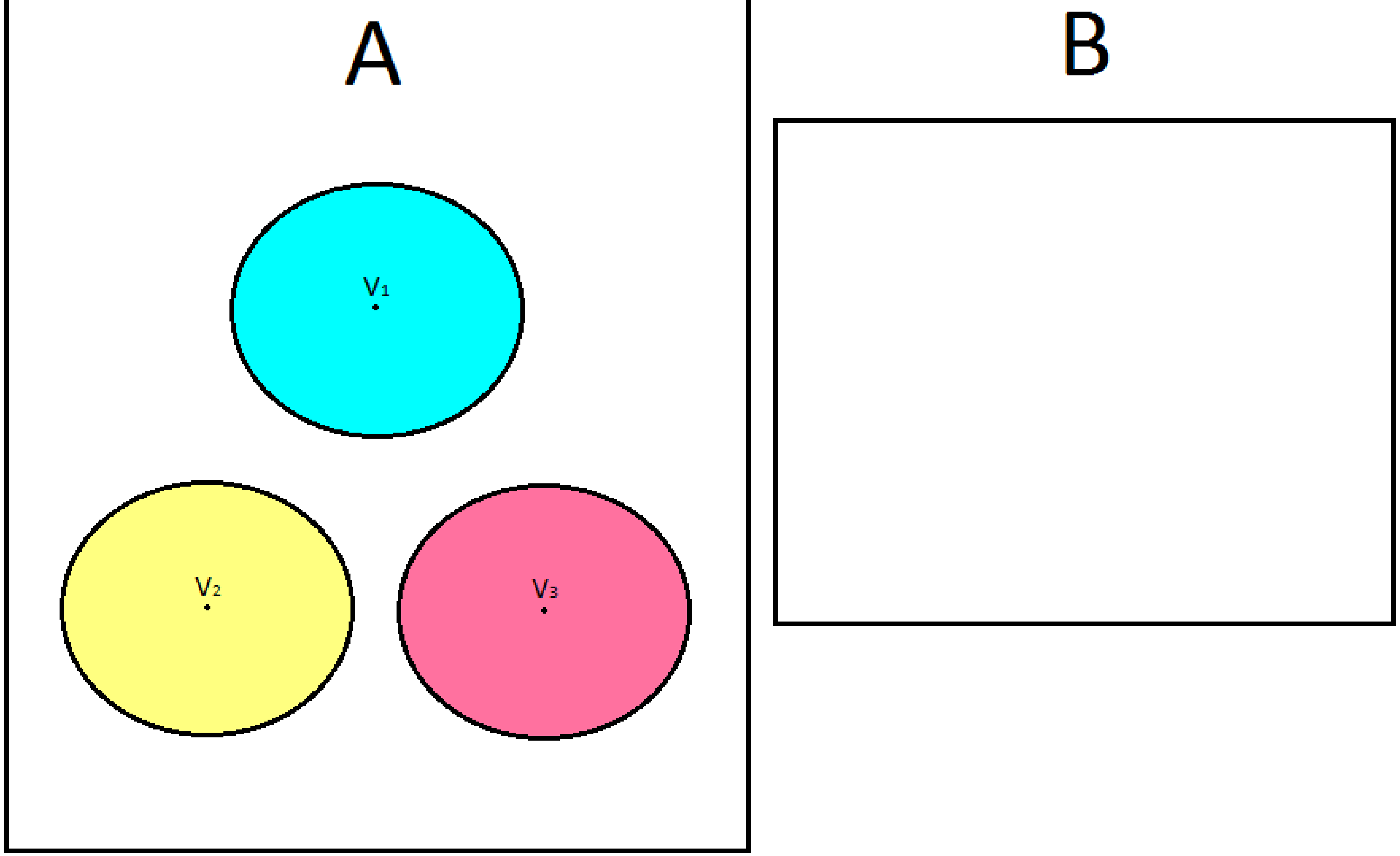
- Theorem (D., A. Ganesh):

$$nW_k - (k-1)\log n \xrightarrow{p} -\infty \quad as\ n \to \infty$$

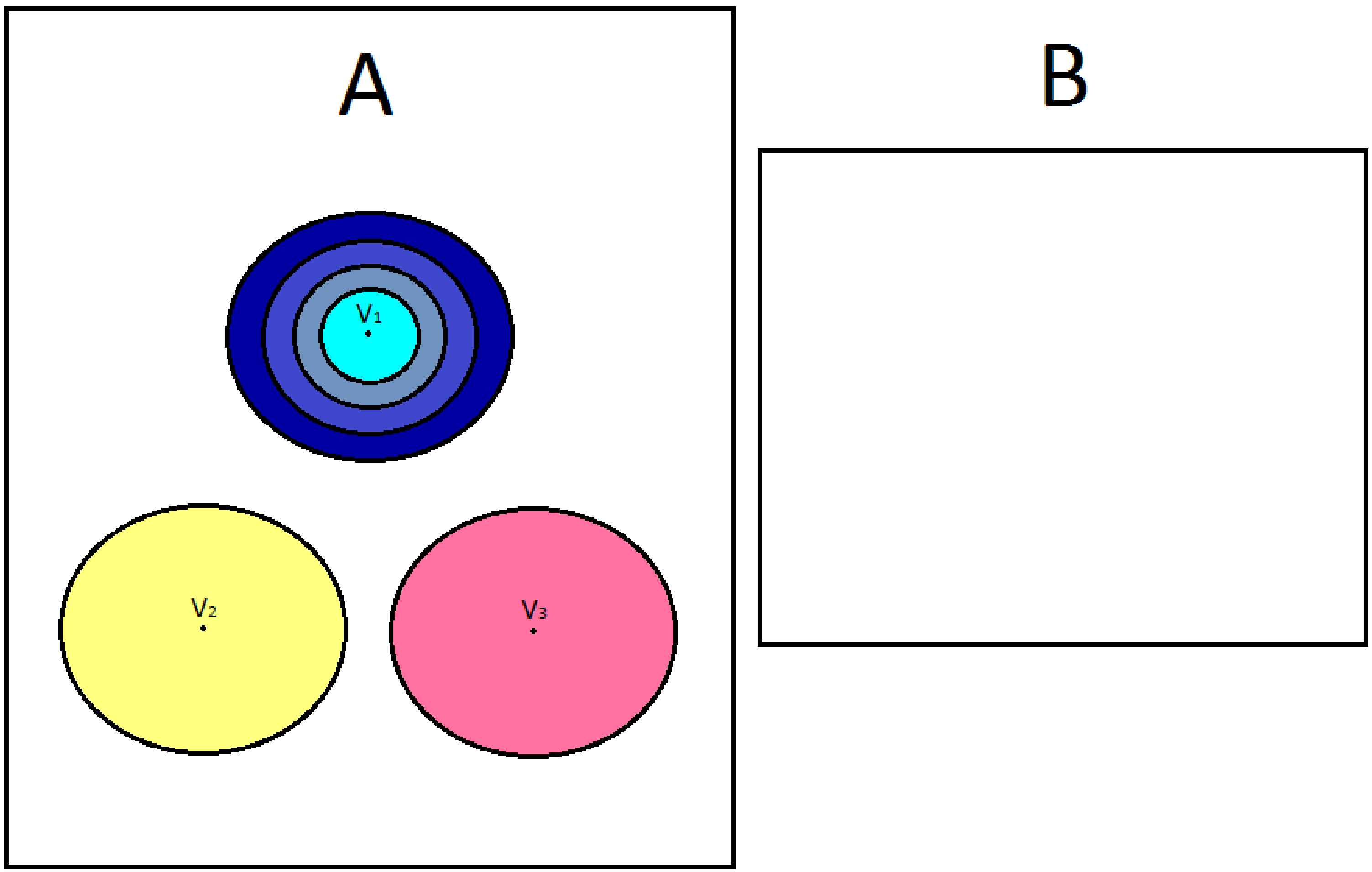
Algorithm



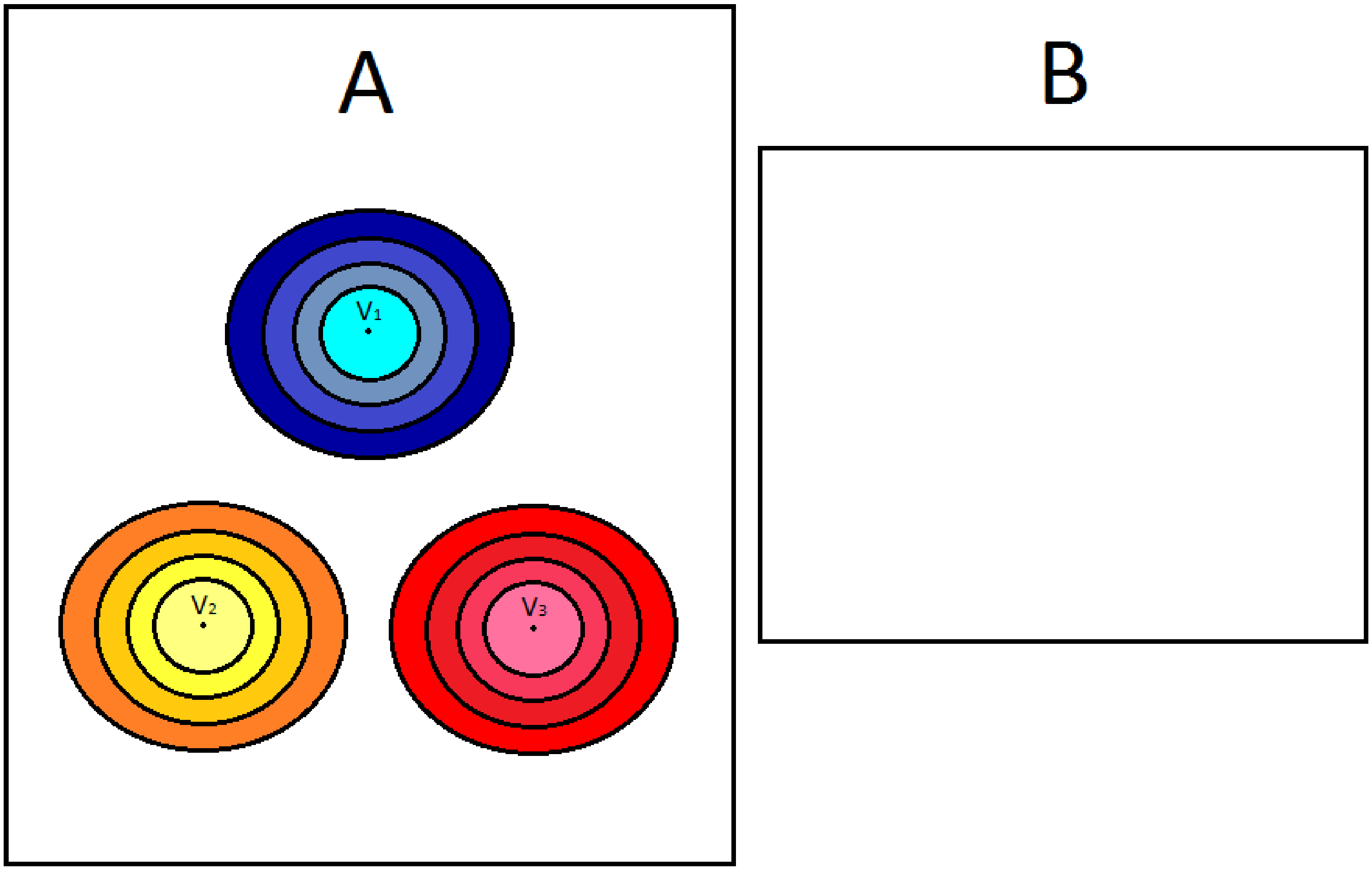
Algorithm



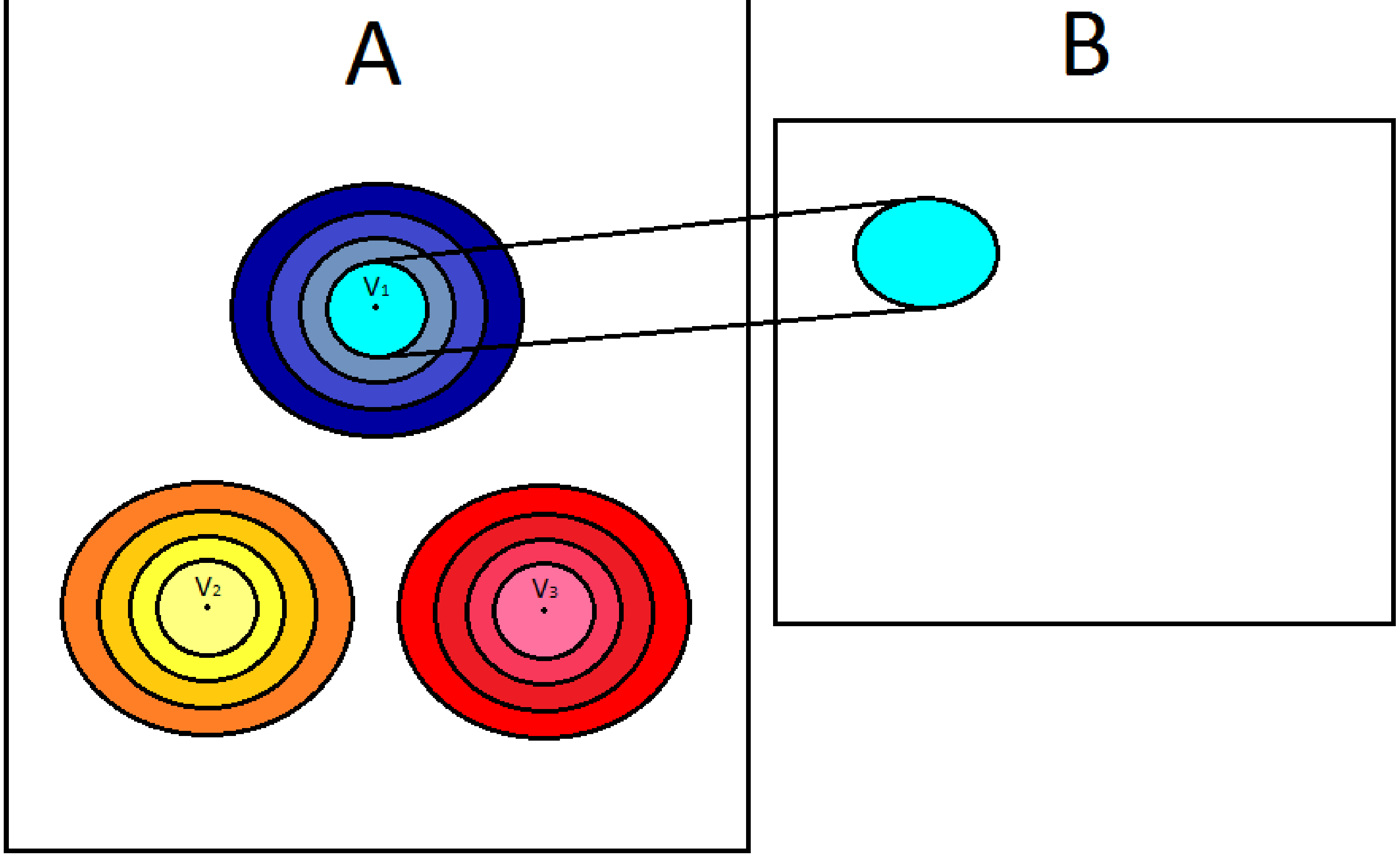
Algorithm



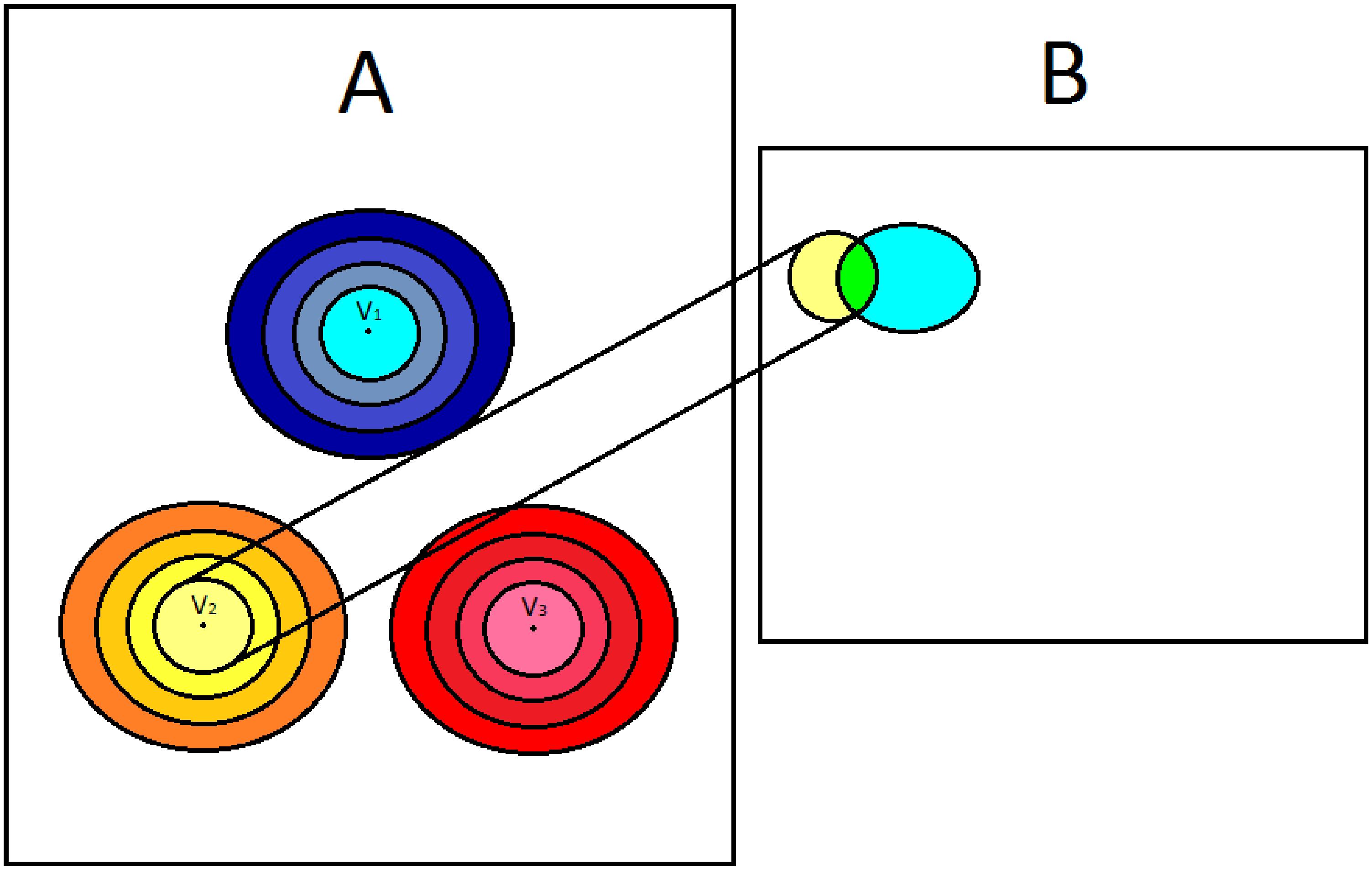
Algorithm



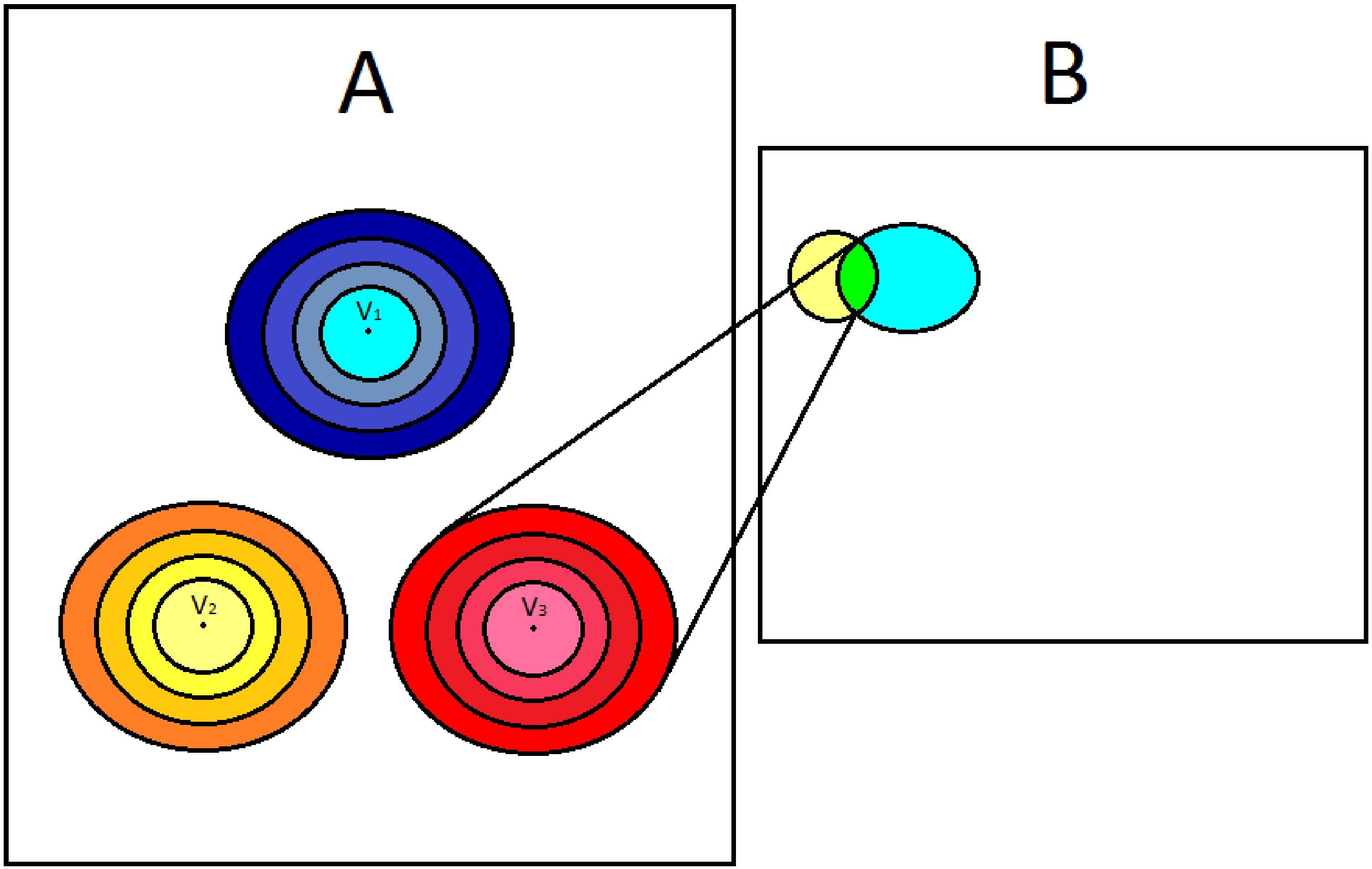
Algorithm



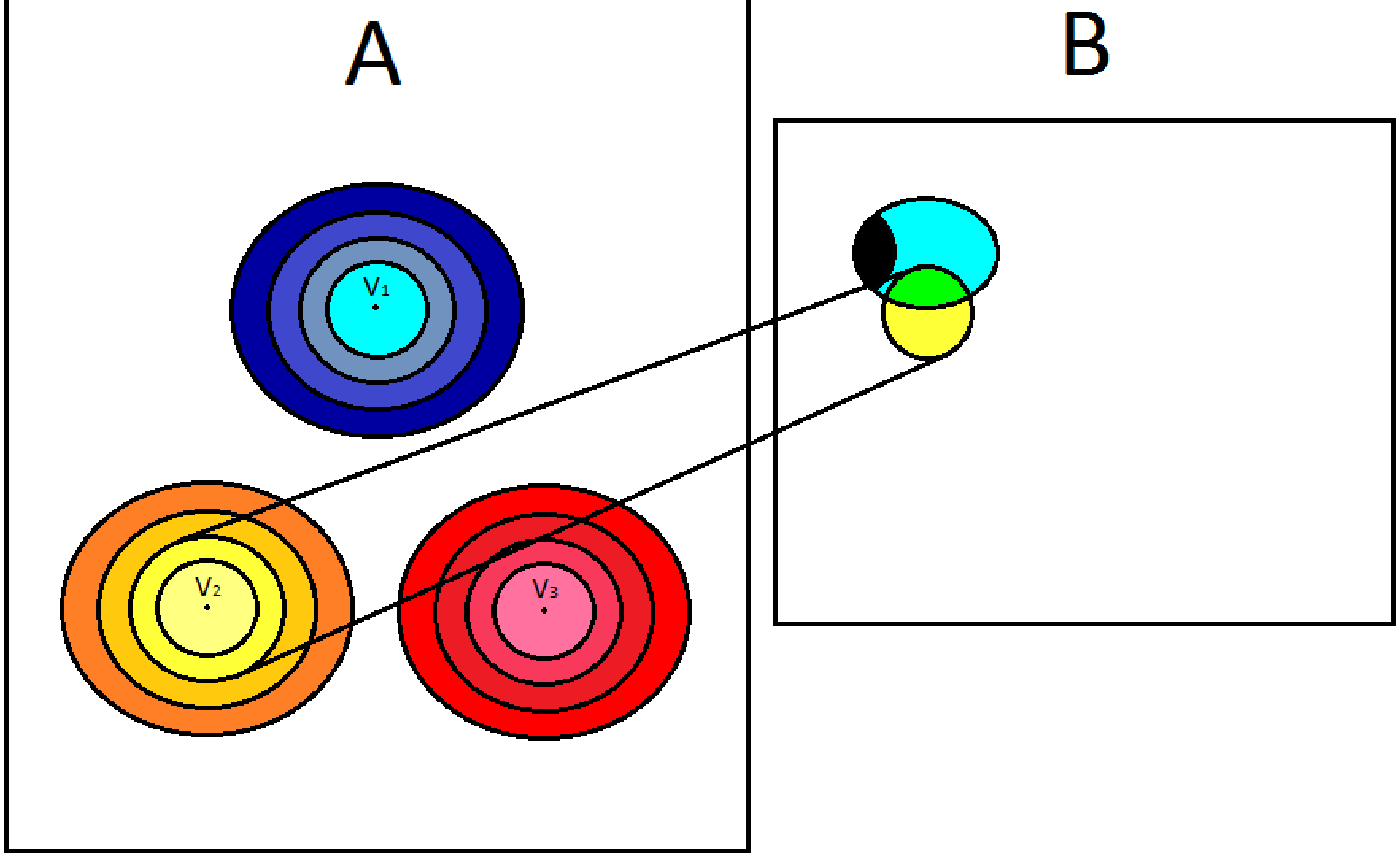
Algorithm



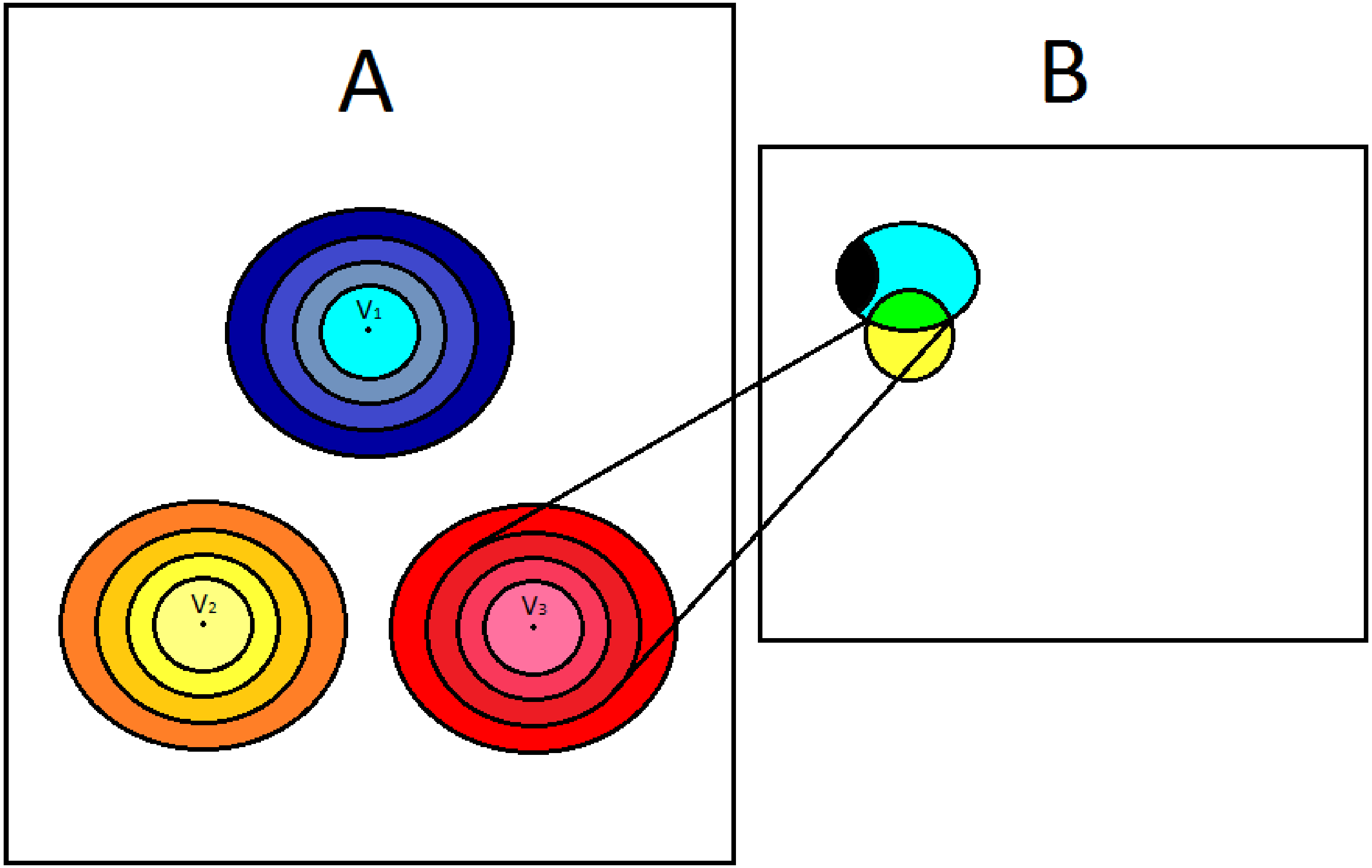
Algorithm



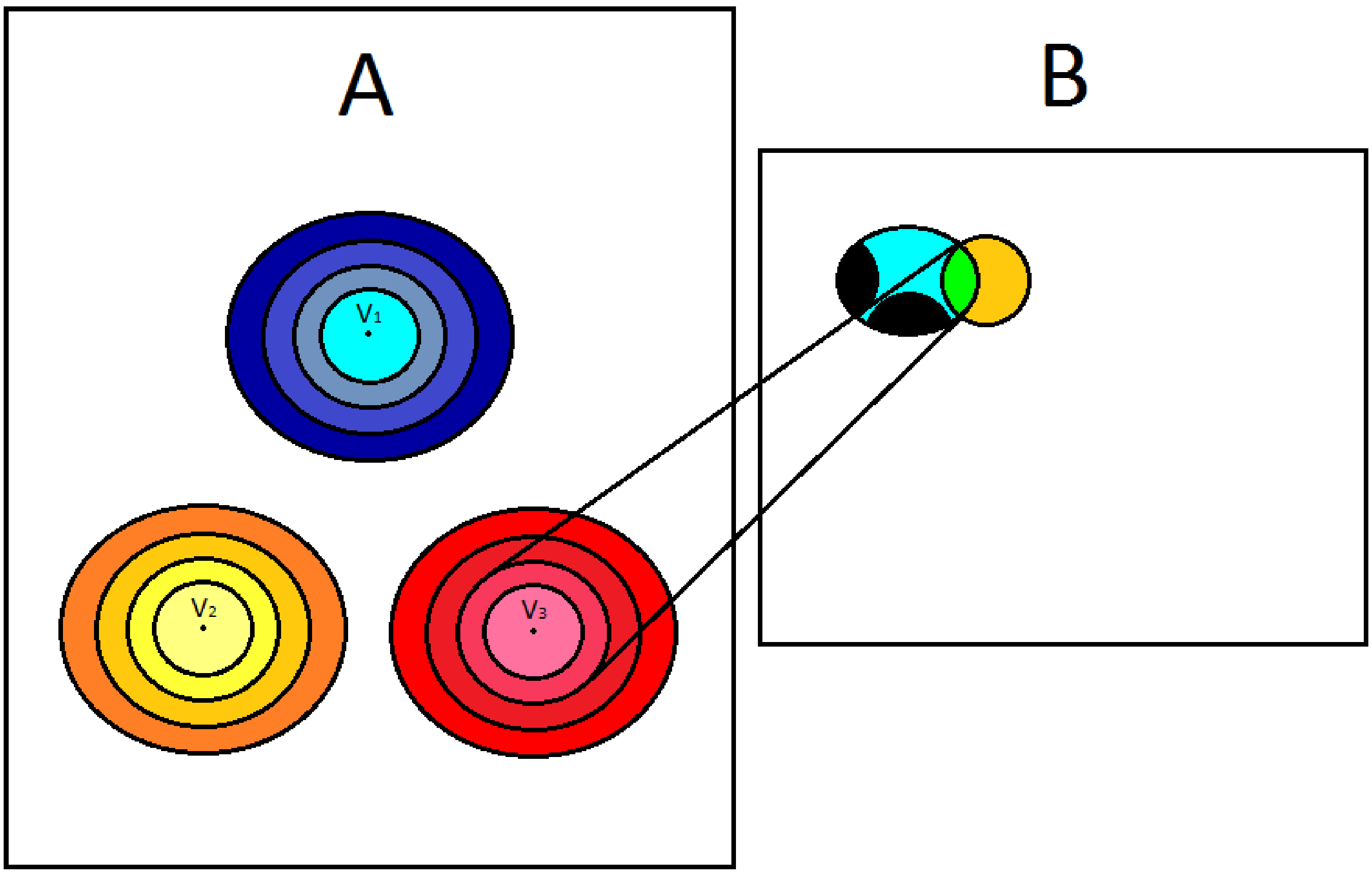
Algorithm



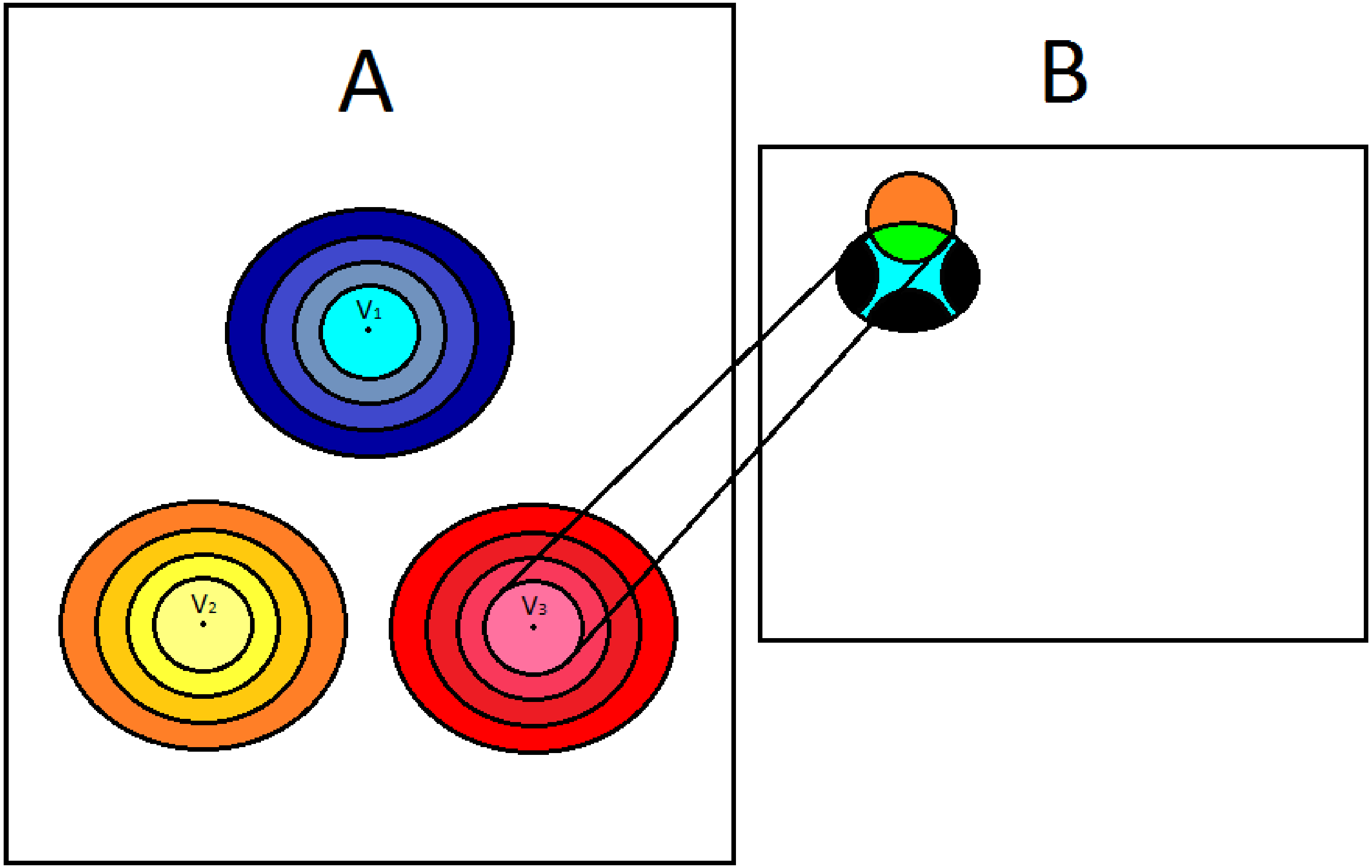
Algorithm



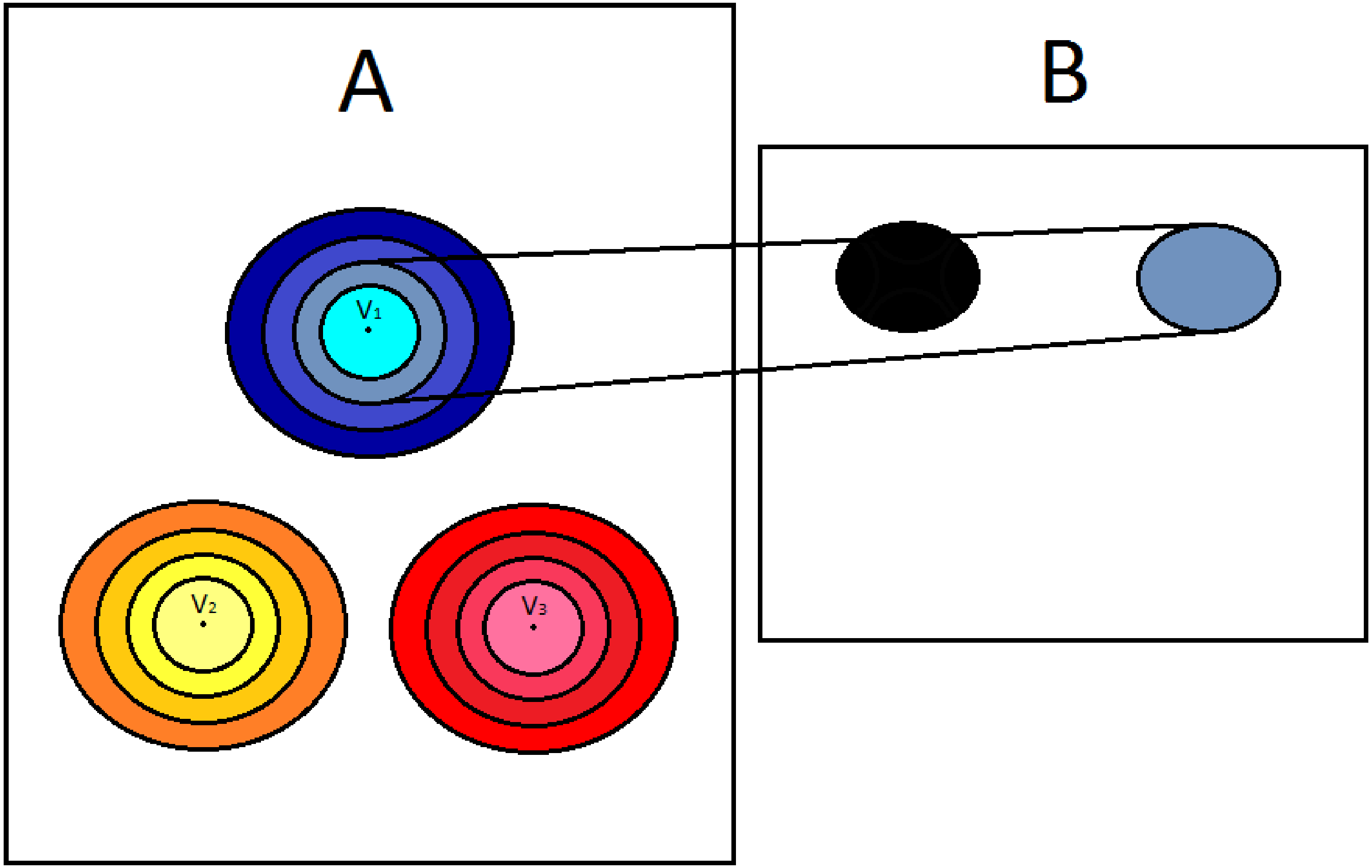
Algorithm



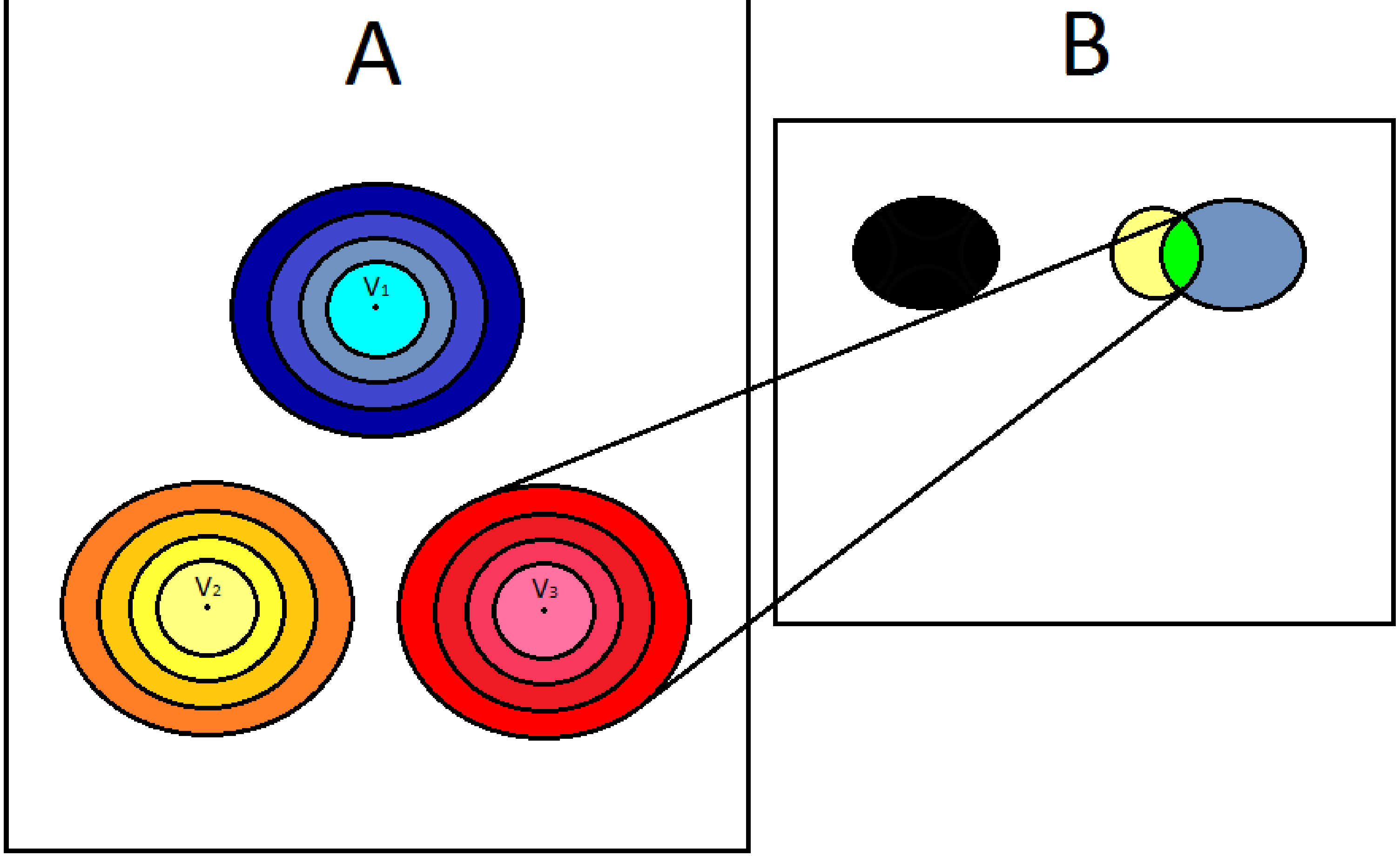
Algorithm



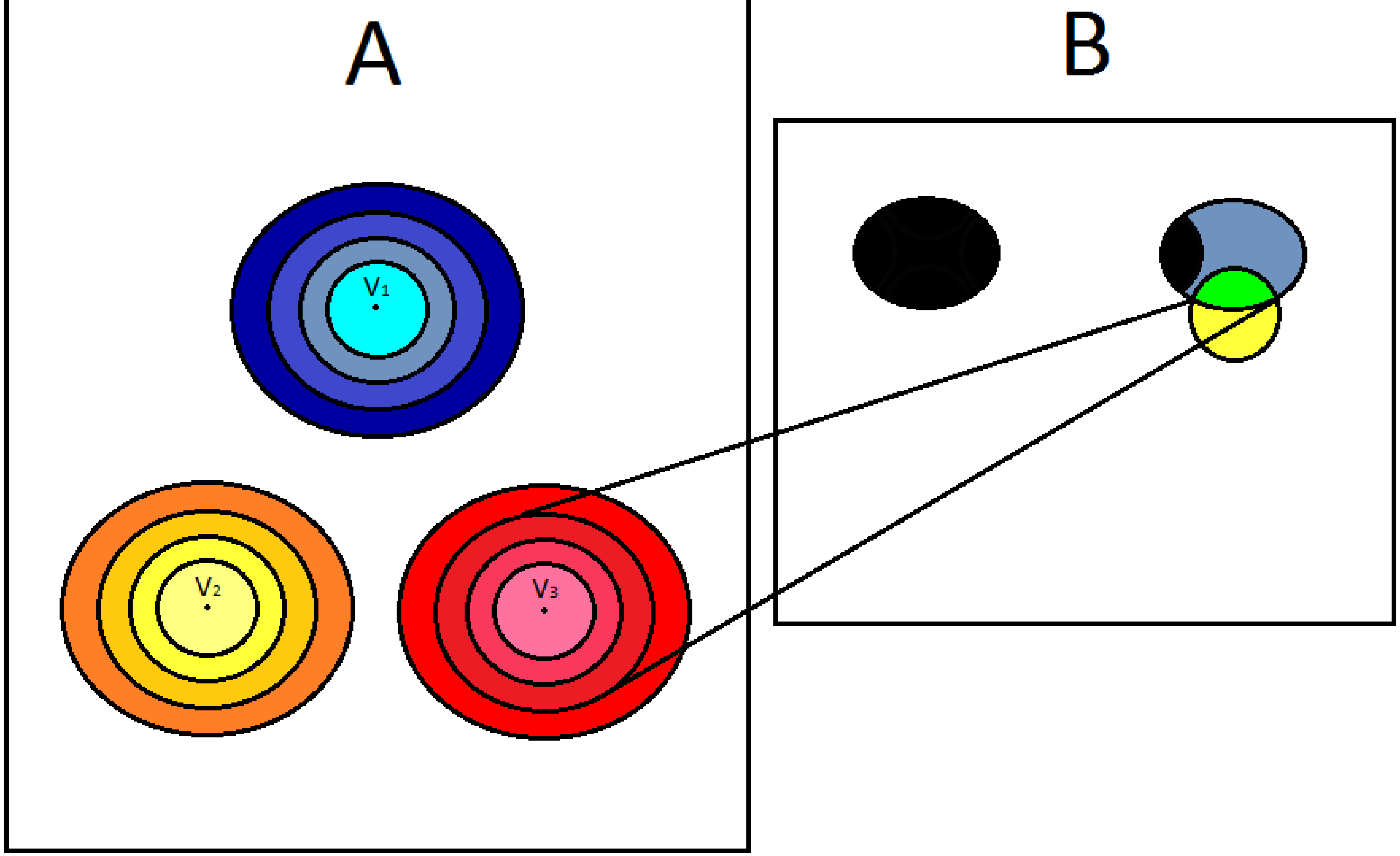
Algorithm



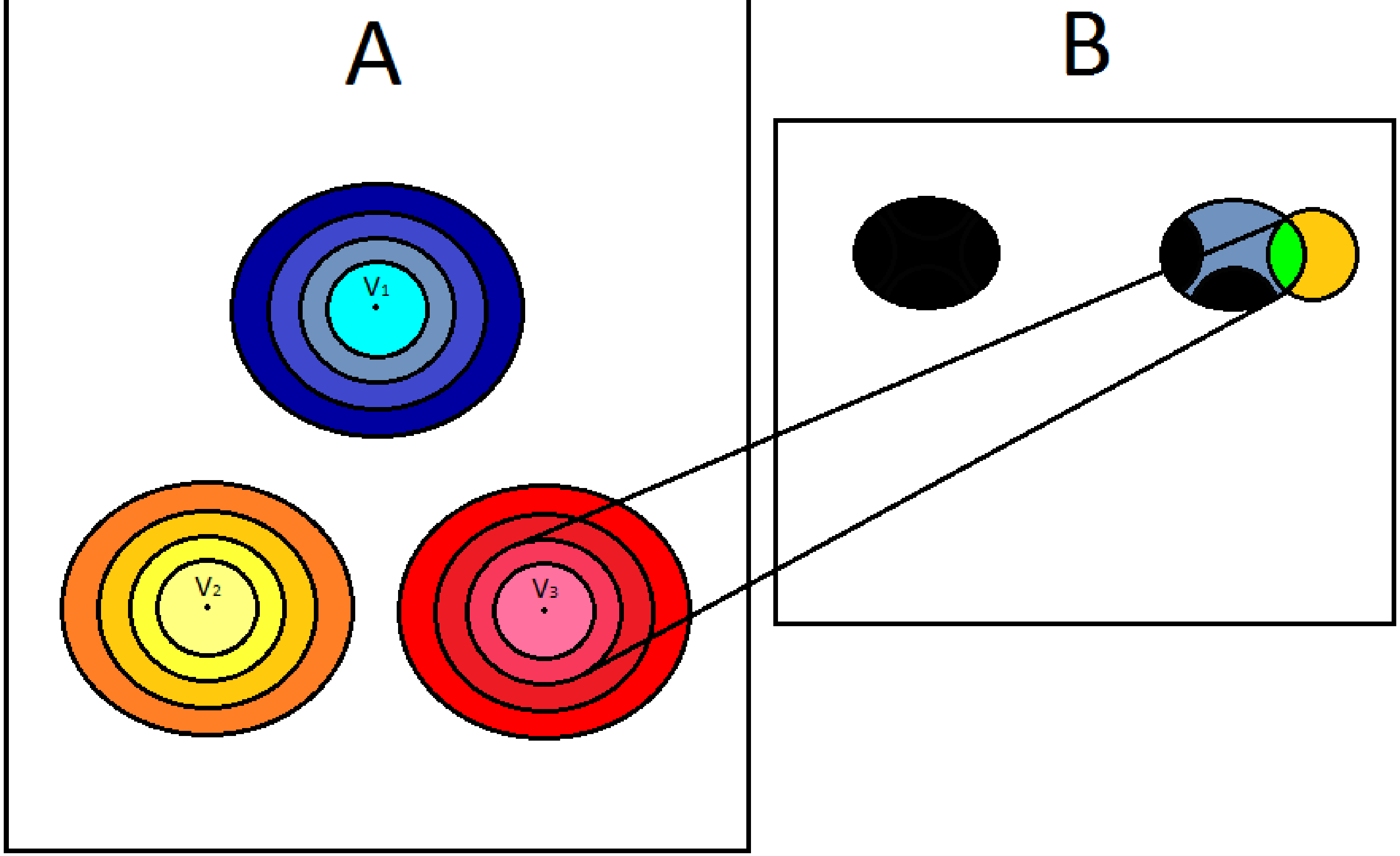
Algorithm



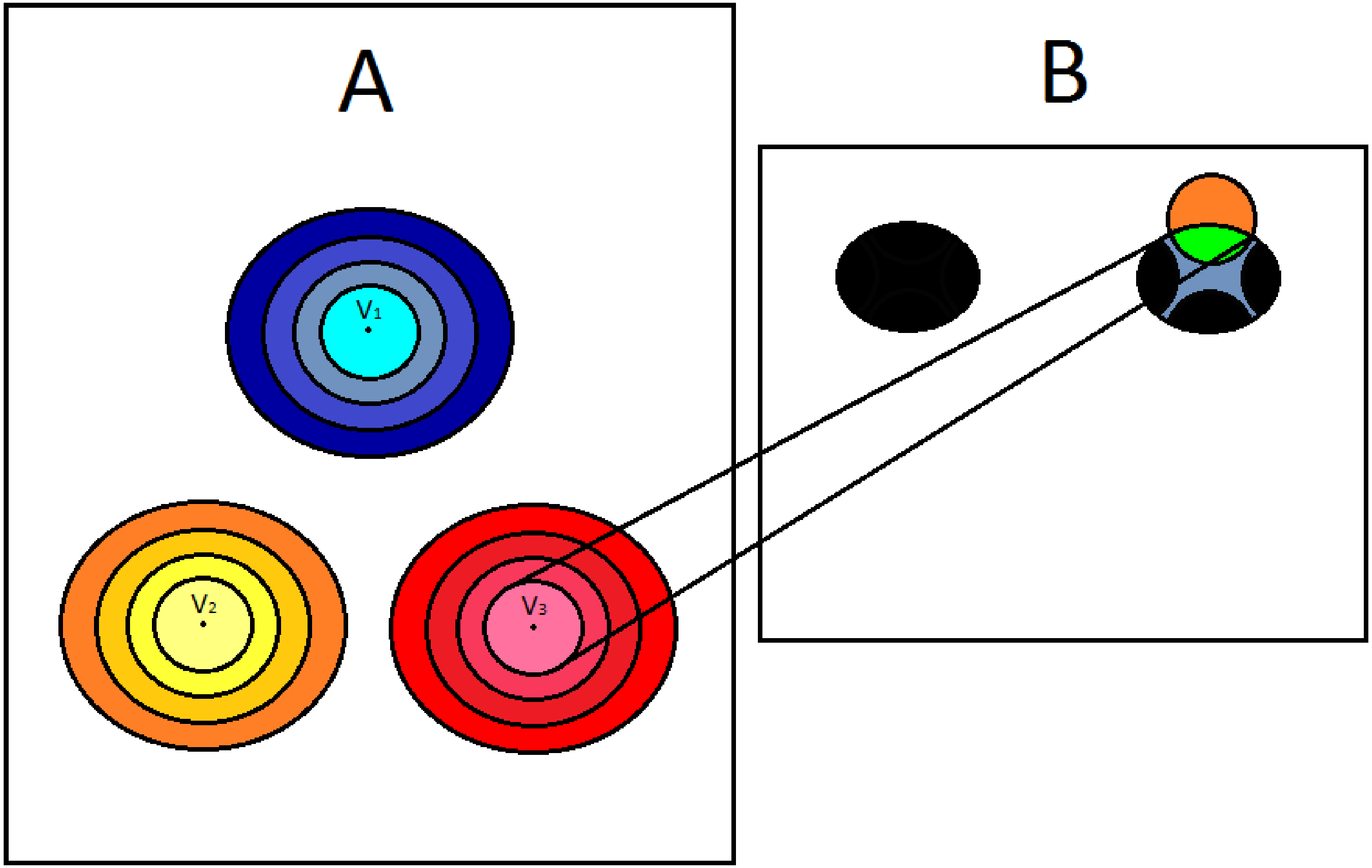
Algorithm



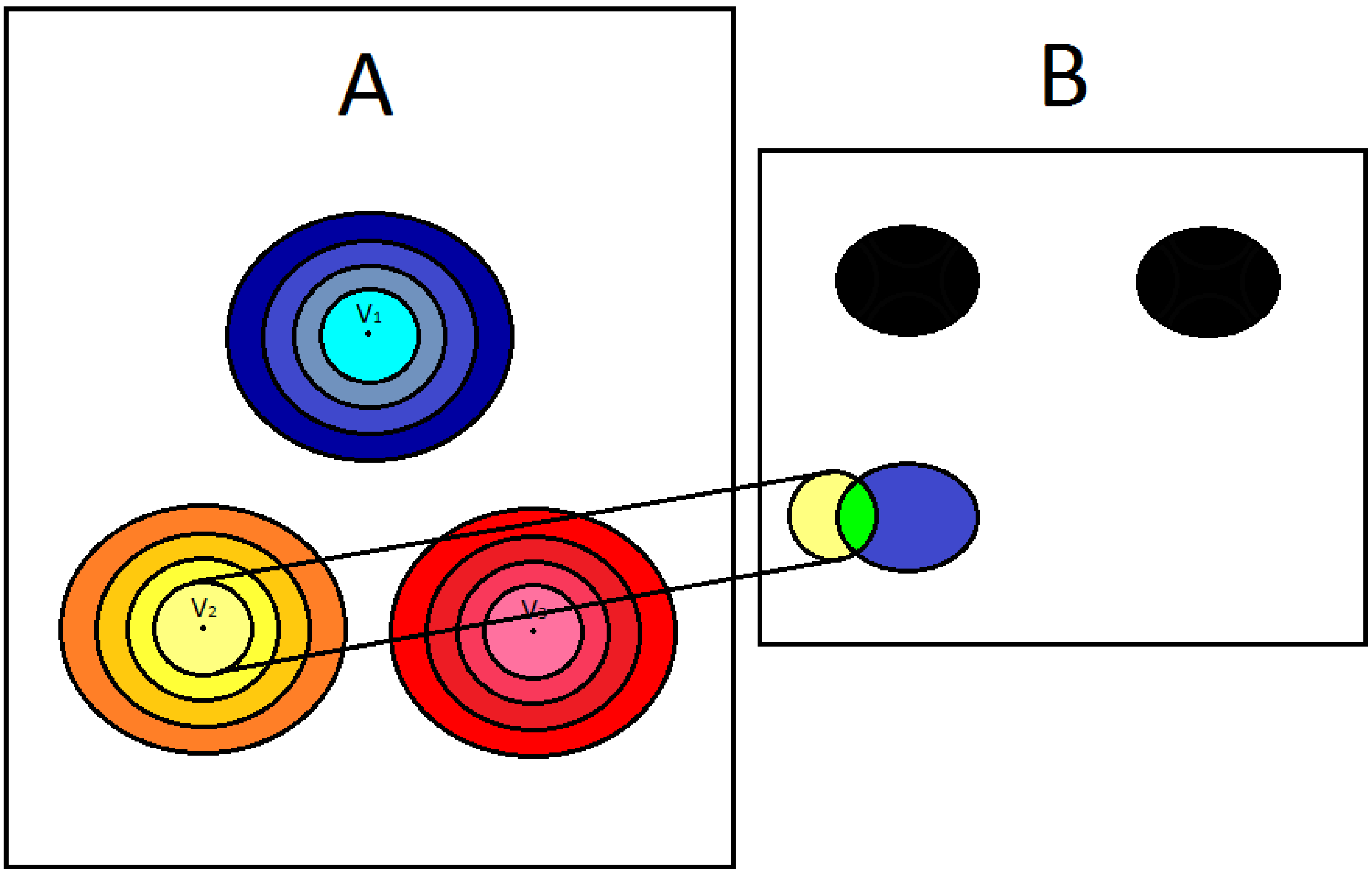
Algorithm



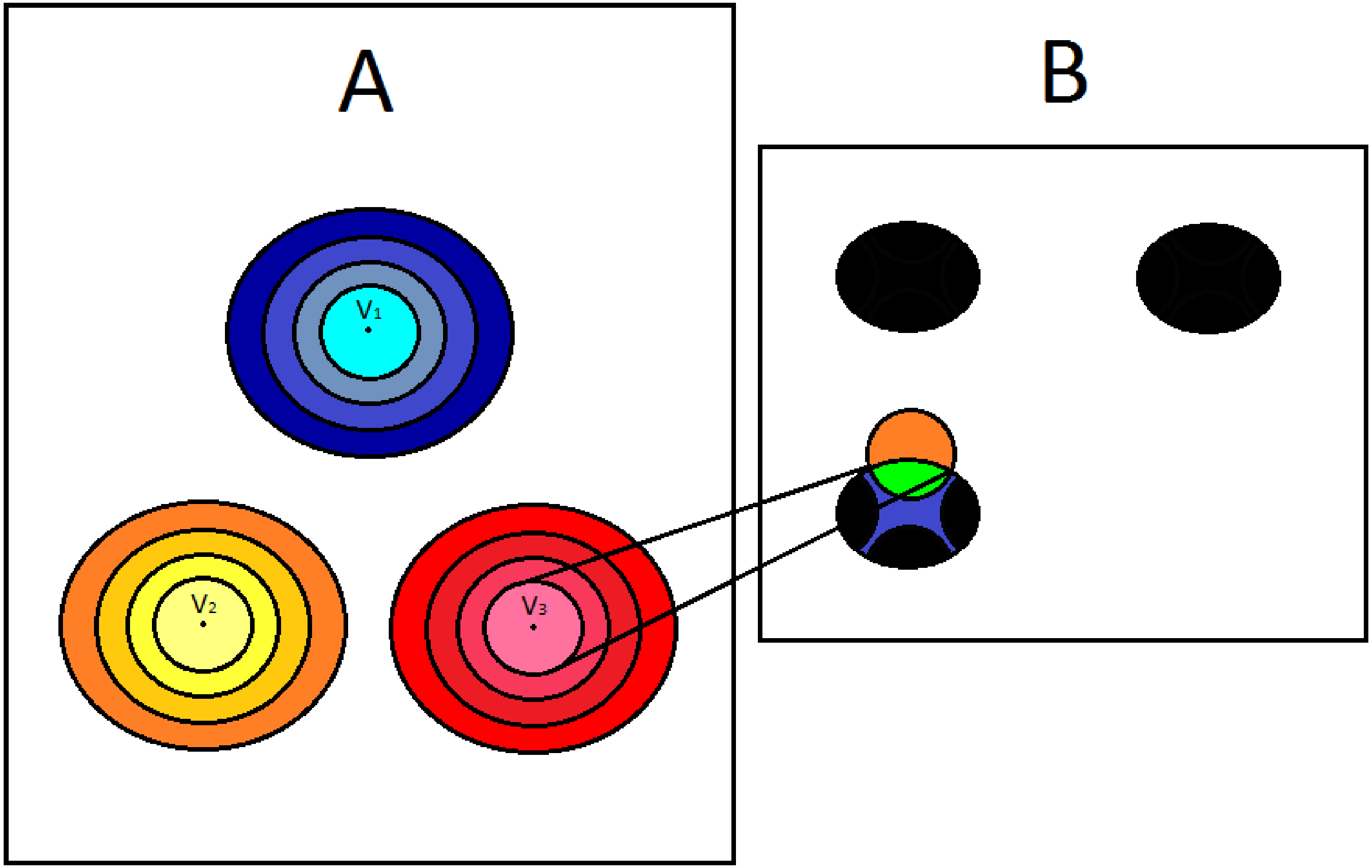
Algorithm



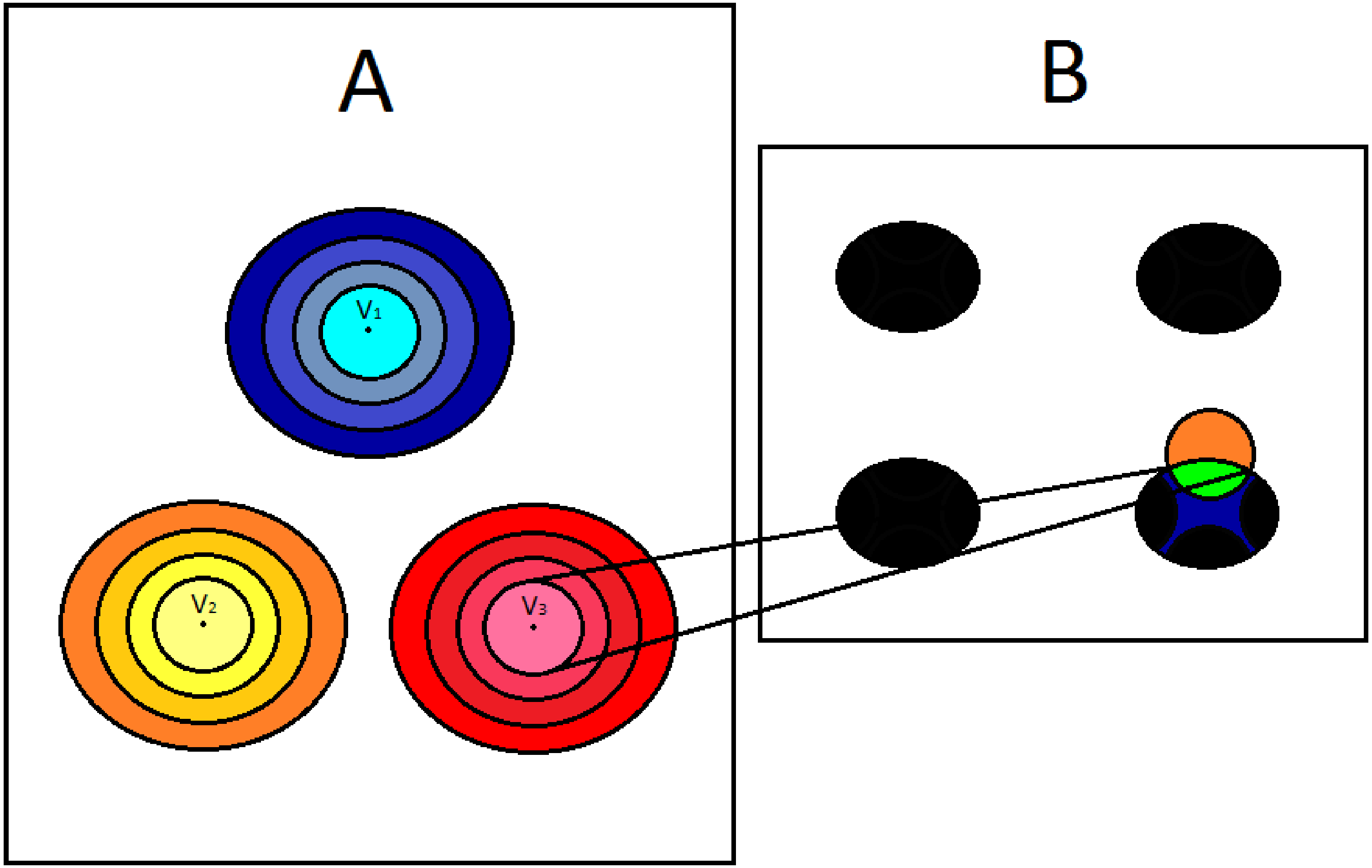
Algorithm



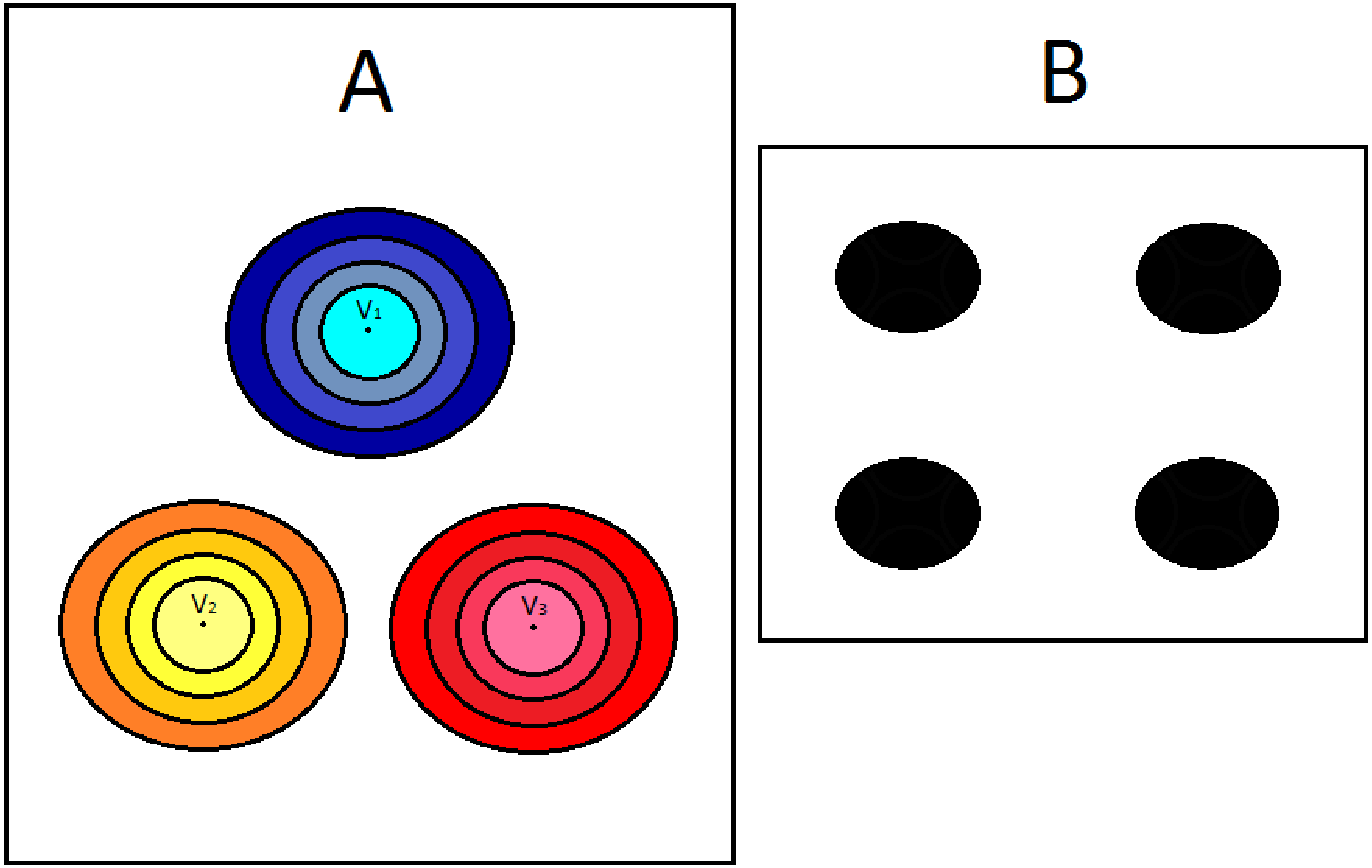
Algorithm



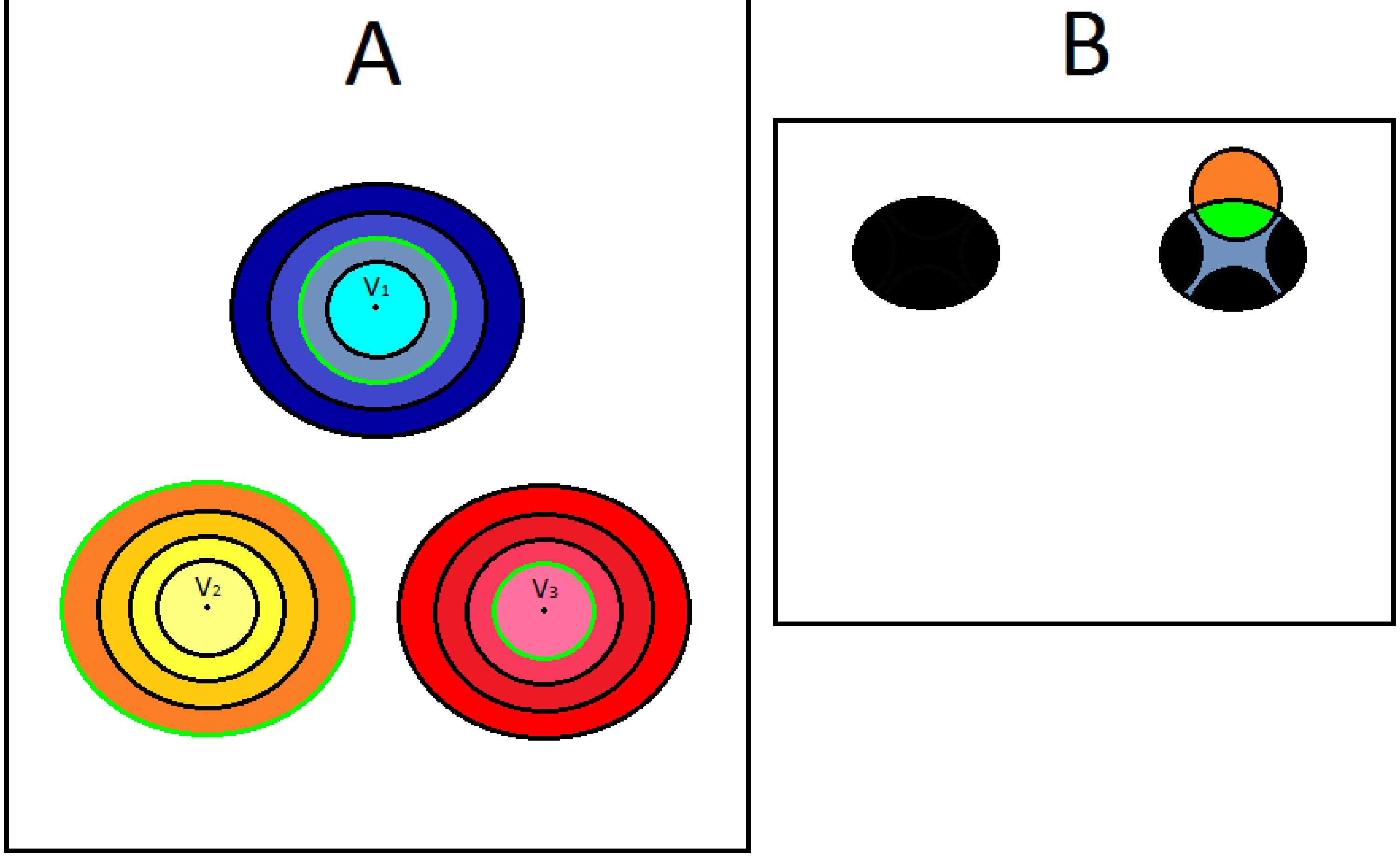
Algorithm



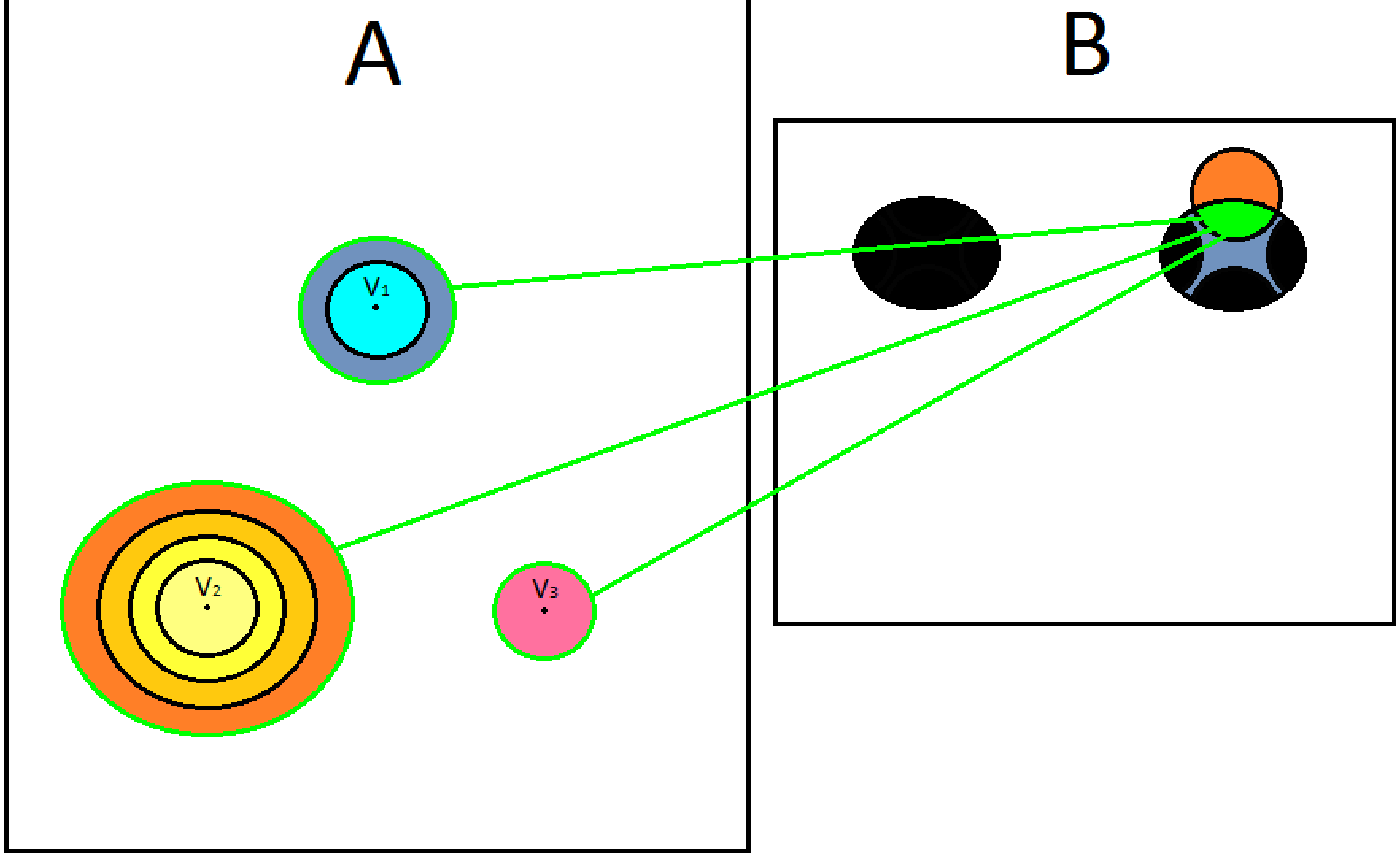
Algorithm



Algorithm



Algorithm



Upper Bound/Lower Bound

- Theorem $\forall \varepsilon > 0, k > 2$:

$$\mathbb{P}\left(-(k-1+\varepsilon) < \frac{nW_k - (k-1)\log n}{\log\log n} < -(k-2-\varepsilon)\right) \to 1 \;\; as\;\, n \to \infty$$

- Conjecture

$$\frac{nW_k - (k-1)\log n}{\log\log n} \xrightarrow{p} -(k-1) \;\; as\;\, n \to \infty$$

Conclusions

- Obtained limiting results for weight of extremal Steiner trees in random edge-weighted graphs
- Made progress towards characterising fluctuations of typical Steiner tree
- Fluctuations progress implies some non-trivial distribution of limiting tree shapes