

1a. Read the MAGIC paper and describe the 4 steps of the MAGIC algorithm by explaining the key mathematical operation applied at each step.

The MAGIC algorithm relies on structure in the data where a diffusion operator is used to learn the underlying lower dimensional manifold for mapping cellular phenotypes. It takes in an observed count matrix and recovered a count matrix that represents the likely expression for each individual cell. First, MAGIC identifies the cells are most similar and aggregates gene expression across these cells to impute gene expression correcting for noise. However, due to sparsity, data diffusion is used to construct a weighted affinity matrix that represents a faithful neighborhood of cells. This is done by first preprocessing the data using PCA. Then, using a Gaussian kernel, the distance matrix is converted to an affinities matrix. The adaptive kernel equalizes the effective number of neighbors for each cell while the non-adaptive kernel collapses the data into the densest regions. Then, from the affinity matrix, the Markov transition matrix can be generated that represents the probability distribution of transitioning from one cell to another in a single step. To reduce the noise, the Markov transition matrix is raised to a power of t , which results in a matrix where each entry represents the probability that a random walk of length t starting at cell i will reach cell j . In the imputation step, MAGIC multiplies the transition matrix by the original data and restores the cells to the underlying manifold. The optimal magnitude of t is selected by considering the impact of t on the final imputed data. The degree of change between the data at time t and time $t-1$ is evaluated and then stopped after this value stabilizes. There is an optimal t which prevents the algorithm from overfitting to the complexities of the data, while also learning the manifold structure and smoothing out the noise.

1b. What concerns do you have about the application of MAGIC to single cell data (if any)? What concerns are mentioned in the paper in the section called 'MAGIC algorithm'?

Some concerns mentioned with the MAGIC algorithm is first, where because of technical noise, the ability to distinguish between similarity due to biological correspondence versus spurious chance is not possible. If a fraction of the transcripts were randomly subsampled, the expression observed across identical cells can appear dissimilar, though the cells can share many neighbors, whereas spurious edges connect cells that share few neighbors. Additionally, MAGIC assumes that gene-gene relationships exist in a subspace corresponding to low-frequency trends in data and considers noise as high frequency. Thus, low-frequency batch effects are not removed and noisy genes at a high frequency will be smoothed out. As such, it's important to choose a diffusion time parameter that regulates the extent of smoothing performed.

1c. Explain how the authors adjust σ_i in an adaptive fashion for each data point.

MAGIC uses an adaptive Gaussian kernel to equalize the effective number of neighbors for each cell, which diminishes the effect of differences in density. The kernel width σ is not fixed but adapted for each cell. In order to adapt σ_i for each cell, the value of $\sigma(i)$ for each cell is equal to the distance to its nearest neighbor: $\sigma(i) = \text{distance}(i, \text{neighbor}(i, k_a))$. To maximize sensitivity to recover fine structure, the k_a is chosen to be small at possible such that the graph is still connected.

$$2[a] H[P(x)] = - \sum_{x \in X} P(x) \log(P(x))$$

$$H[1/N] = - \sum_{x \in N} \frac{1}{N} \log\left(\frac{1}{N}\right)$$

$$= N \cdot \frac{1}{N} \log N$$

$$= \boxed{\log N}$$

$$[b] P(x) = \begin{cases} x=1, 1 \\ x \neq 1, 0 \end{cases}$$

$$H[P(x)] = -(1 \cdot \log(1) + \sum_{x \neq 1} 0 \cdot \log(0))$$

Since $0 \log(0) = 0$ asymptotically,

$$H[P(x)] = -(1 \cdot \log(1) + 0)$$

$$= \boxed{-\log(1)} = 0$$

$$= -\log(1) = \boxed{0}$$

[c] Entropy measures the uncertainty in a probability distribution.

The distribution in part a was uniform, which maximizes the entropy because of equiprobable outcomes. For part b, we had a discrete probability distribution, where both events have a certain outcome (either 0 or 1 depending on x). Thus, the entropy is equal to 0.