Interactive Visual Analysis of Single-Cell Data with HSNE

Thomas Höllt, Vincent van Unen, Nicola Pezzotti, Na Li, Marcel J.T. Reinders, Elmar Eisemann, Frits Koning, Anna Vilanova, and Boudewijn P.F. Lelieveldt
t-SNE

- Non-linear dimensionality reduction
  - Compute local neighborhoods in hi-D
  - Model low-D to preserve neighborhoods
    - Preserves high-D clusters!
- Computationally intensive
- Crowding problem
t-SNE

• Non-linear dimensionality reduction
  • Compute local neighborhoods in hi-D
  • Model low-D to preserve neighborhoods
    ➡ Preserves high-D clusters!

• Computationally intensive
• Crowding problem
t-SNE

• Non-linear dimensionality reduction
  • Compute local neighborhoods in hi-D
  • Model low-D to preserve neighborhoods
    ∆ Preserves high-D clusters!

• Computationally intensive

• Crowding problem
t-SNE

- Non-linear dimensionality reduction
  - Compute local neighborhoods in hi-D
  - Model low-D to preserve neighborhoods
    - Preserves high-D clusters!
- Computationally intensive
- Crowding problem
Crowding

1 million cells
16GB - 19.5 h

2 million cells
38GB - 45.5 h

5 million cells
99GB - 252 h
Crowding

1 million cells
16GB - 19.5 h

2 million cells
38GB - 45.5 h

5 million cells
99GB - 252 h
HSNE
HSNE

marker c → marker b
HSNE (2 levels)

HSNE 1

HSNE 2

Overview Level

Aol (# Events)
HSNE

marker c

marker b

HSNE (2 levels)

marker a

HSNE 1

HSNE 2

Aol
(# Events)

Overview Level

Data Level
HSNE

Overview Level

Data Level

Aoi
(# Events)

HSNE 1

HSNE 2

HSNE 1

HSNE 2

marker a

marker b

marker c
HSNE

marker c

marker b

HSNE (2 levels)

marker a

HSNE 1

HSNE 2

Overview Level

Data Level

Aoi
(# Events)
HSNE

HSNE (2 levels)

Overview Level

Data Level

HSNE - Construction

• Aggregate cells by *Landmarks*
• Landmarks selected based on local connectivity in neighborhood graph
• Cells represented by Landmark define Area of Influence (AoI)
• Similarity (above data level) defined by overlap AoI
• Retains non-linearity throughout the hierarchy
Interactive Exploration
CD4+ T cells in blood and intestine

- Single-cell mass cytometry dataset on gastrointestinal diseases
- 5M cells, 40 dimensions (simultaneously measure proteins)
- Less than 30 minute computation time
CD7  CD3  CD4  TCRγδ  CD11c

CD8a  CD19

CD45RA  CCR7  CD28  CD161

CD127  CD38
CD4+ T cells in blood and intestine
CD4+ T cells in blood and intestine
CD4⁺ T cells in blood and intestine
http://www.cytosplore.org
http://www.lcbc.nl

Acknowledgements:
STW / NWO Grant 12720 VAnPIRe
LKEB, IHB, LCBC @ LUMC
CGV @ TU Delft