# Genomic relatedness strengthens genetic connectedness across management units

Haipeng Yu

Annual Meeting of the American Society of Animal Science

July 11, 2017

# Introduction

**"Genetic Connectedness"**

- In animal breeding, it refers to a measure of genetic relatedness across management units (e.g., herds and flocks).
- Best linear unbiased prediction (BLUP): genetic links or relatedness

**Objective:**

- Assess measures of genetic connectedness across management units with use of genomic information.

# Prediction Error Variance (PEV)

**Inverse of MME coefficient matrix:**

$$
\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{K}^{-1}\lambda \end{bmatrix}^{-1}
$$
$$
= \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}
$$

$$
\mathrm{PEV}(\hat{u}) = (\mathbf{Z'MZ} + \mathbf{K}^{-1}\lambda)^{-1}\sigma_{\mathbf{e}}^{2}
$$
$$
= \mathbf{C}^{22}\sigma_{e}^{2}
$$

where $\mathbf{M}$ is the absorption matrix given by

$$
\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'}
$$

## Statistics of Pairwise Genetic Connectedness

Prediction error variance of the difference ($\textbf{PEVD}_{ij}$)

$$
\begin{aligned}
\text{PEVD}(\hat{u}_i - \hat{u}_j) &= [\text{PEV}(\hat{u}_i) + \text{PEV}(\hat{u}_j) - 2\text{PEC}(\hat{u}_i, \hat{u}_j)] \\
&= (\textbf{C}_{ii}^{22} - \textbf{C}_{ji}^{22} - \textbf{C}_{ij}^{22} + \textbf{C}_{jj}^{22})\sigma_\epsilon^2 \\
&= (\textbf{C}_{ii}^{22} + \textbf{C}_{jj}^{22} - 2\textbf{C}_{ij}^{22})\sigma_\epsilon^2,
\end{aligned}
$$

Coefficient of determination ($\textbf{CD}_{ij}$)

$$
\text{CD}_{ij} = 1 - \lambda \frac{\textbf{C}_{ii}^{22} + \textbf{C}_{jj}^{22} - 2\textbf{C}_{ij}^{22}}{\textbf{K}_{ii} + \textbf{K}_{jj} - 2\textbf{K}_{ij}}.
$$

Prediction error correlation ($\textbf{r}_{ij}$)

$$
r_{ij} = \frac{PEC(\hat{u}_i, \hat{u}_j)}{\sqrt{PEV(\hat{u}_i)PEV(\hat{u}_j)}}.
$$

# Choice of K

- **Numerator Relationship Matrix (A)** : $a_{ii} = 1 + F_i$; $a_{ij}$ is equal to twice the kinship coefficient between animal $i$ and $j$

- **Genomic Relationship Matrix (G)** : $G = \frac{ZZ'}{m}$, where $Z_{\cdot j} = \frac{z_{\cdot j} - 2p_j}{\sqrt{2p_j(1-p_j)}}$

- **$G_{0.5}$** : $MAF(p_j)$ in based generation is treated as 0.5

- **Scaled G matrix (Gs)** : $G_{s_{ij}} = \frac{(G_{s_{max}} - G_{s_{min}})(G_{ij} - G_{min})}{G_{max} - G_{min}}$

- **H matrix**:

$$\begin{bmatrix} A_{11} + A_{12}G{A_{22}}^{-1}(G - A_{22}){A_{22}}^{-1}A_{21} & A_{12}{A_{22}}^{-1}G \\ G{A_{22}}^{-1}A_{21} & G \end{bmatrix}$$

# Materials

**Mice Data:**

- public real data from Wellcome Trust Centre for Human Genetics
- $1,814$ individuals, $10,346$ markers and $523$ cages

**Cattle Data:**

- real data coupled with simulated management units , publicly available in the synbreed R package
- $500$ individuals and $7,250$ markers
- simulated management units from $8$ clusters (K-medoid and average silhouette width algorithm)
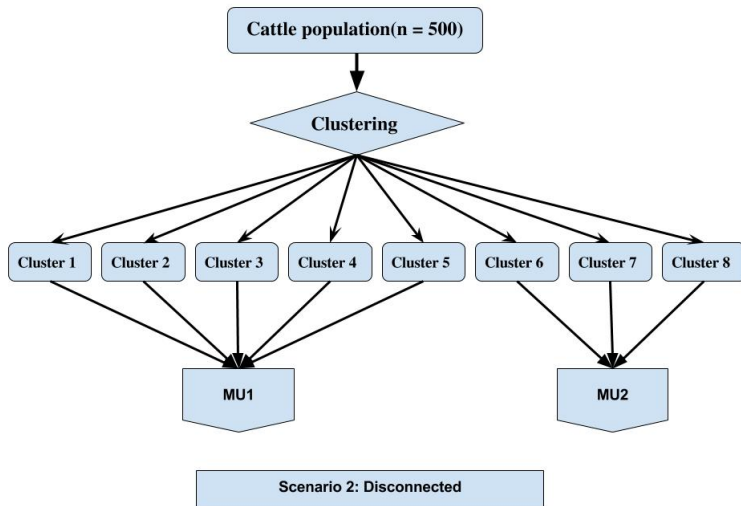
# Average genetic connectedness measures across management units in the mice data

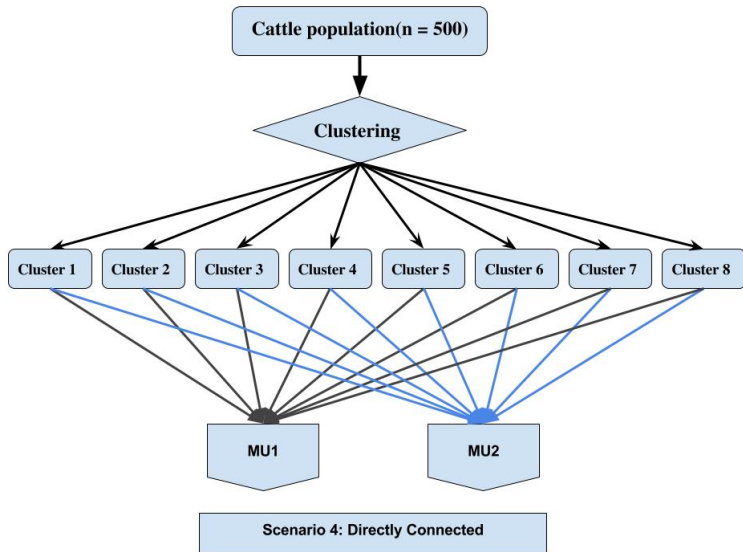| Methods | Kernels | Heritability ($h^2$) | |
| --- | --- | --- | --- |
| | | 0.8 | 0.2 |
| | **A** | 1.299 | 1.331 |
| PEVD | **G** | 0.456 | 1.037 |
| | $\mathbf{G_{0.5}}$ | 0.374 | 0.824 |
| | $\mathbf{G_s}$ | 0.532 | 1.254 |
| | **A** | 0.034 | 0.009 |
| CD | **G** | 0.662 | 0.234 |
| | $\mathbf{G_{0.5}}$ | 0.640 | 0.207 |
| | $\mathbf{G_s}$ | 0.690 | 0.270 |
| | **A** | 0.004 | 0.003 |
| $r_{ij}$ | **G** | -0.001 | -0.001 |
| | $\mathbf{G_{0.5}}$ | 0.559 | 0.433 |
| | $\mathbf{G_s}$ | 0.496 | 0.270 |

# Cattle Management Units Simulation -Scenario 1
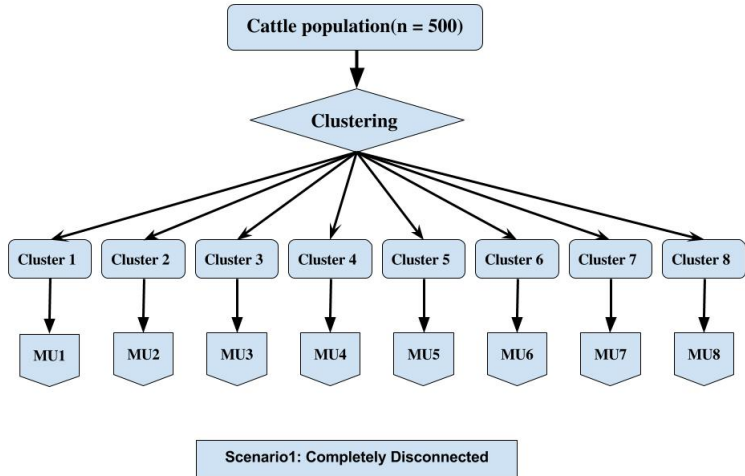
# Cattle Management Units Simulation -Scenario 2

# Cattle Management Units Simulation -Scenario 3
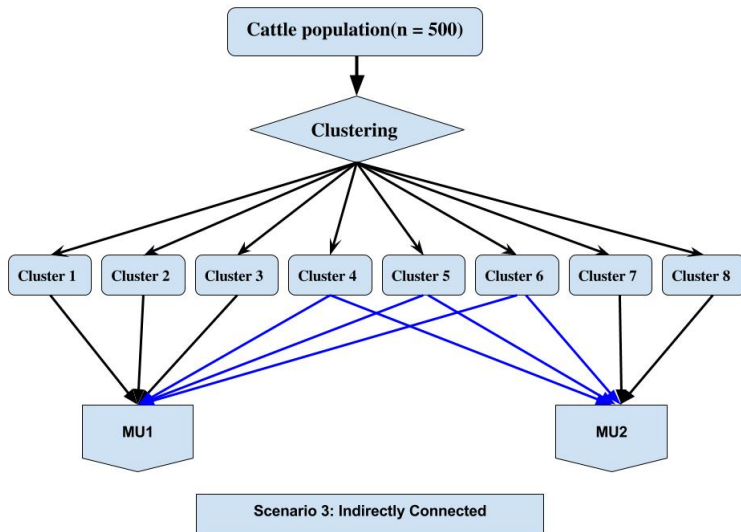
# Cattle Management Units Simulation -Scenario 4

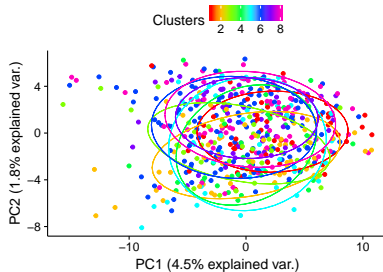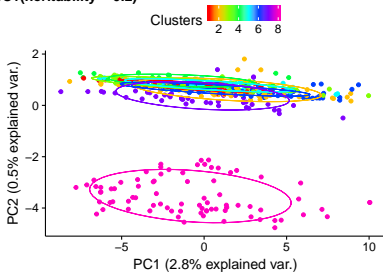# Average genetic connectedness across management units for cattle data

| Scenarios | Methods | Kernels | Heritability ($h^2$) | |
|---|---|---|---|---|
| | | | 0.8 | 0.2 |
| S1 | PEVD | **A** | 0.077 | 0.102 |
| | | **G** | 0.051 | 0.085 |
| | | **$G_{0.5}$ ($G_s$)** | 0.039 (0.066) | 0.066 (0.110) |
| | CD | **A** | 0.324 | 0.112 |
| | | **G** | 0.539 | 0.224 |
| | | **$G_{0.5}$ ($G_s$)** | 0.528 (0.558) | 0.195 (0.265) |
| | $r_{ij}$ | **A** | 0.017 | 0.005 |
| | | **G** | -0.014 | -0.007 |
| | | **$G_{0.5}$ ($G_s$)** | 0.725 (0.468) | 0.465 (0.174) |
| S4 | PEVD | **A** | 0.007 | 0.007 |
| | | **G** | 0.005 | 0.007 |
| | | **$G_{0.5}$ ($G_s$)** | 0.004 (0.007) | 0.005 (0.009) |
| | CD | **A** | 0.125 | 0.048 |
| | | **G** | 0.367 | 0.132 |
| | | **$G_{0.5}$ ($G_s$)** | 0.362 (0.384) | 0.114 (0.158) |
| | $r_{ij}$ | **A** | 0.024 | 0.008 |
| | | **G** | -0.007 | -0.002 |
| | | **$G_{0.5}$ ($G_s$)** | 0.741 (0.502) | 0.470 (0.181) |

# Principle Component Analysis for CD
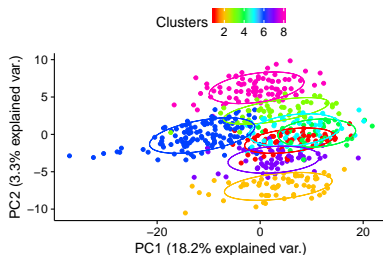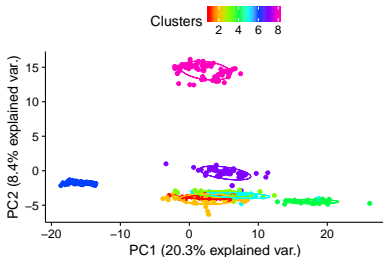
# Principle Component Analysis for CD
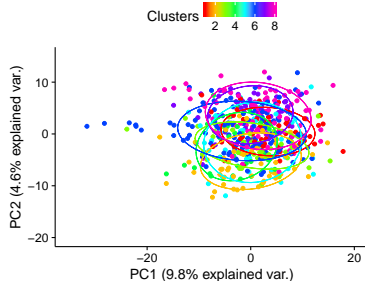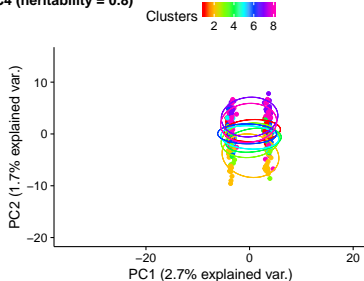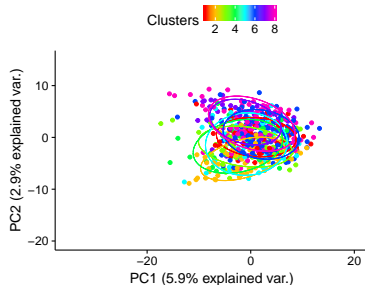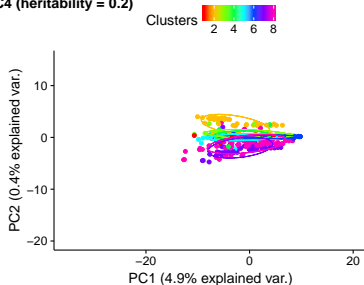
# Conclusion

- Genomic relatedness strengthens measures of genetic connectedness across management units compared to use of pedigree information.

- Coefficient of determination (CD) yields consistent measures of connectedness.

- Genomics provides relationships between units that appear disconnected according to the pedigree (independence assumption at baseline population)

# Acknowledgement

- Dr. Ron Lewis

- Dr. Matthew Spangler

- Dr. Dale Van Vleck

- Dr. Larry Kuehn

- Dr. Gota Morota