Noisy Circle: End-to-End gflow Workflow

Overview

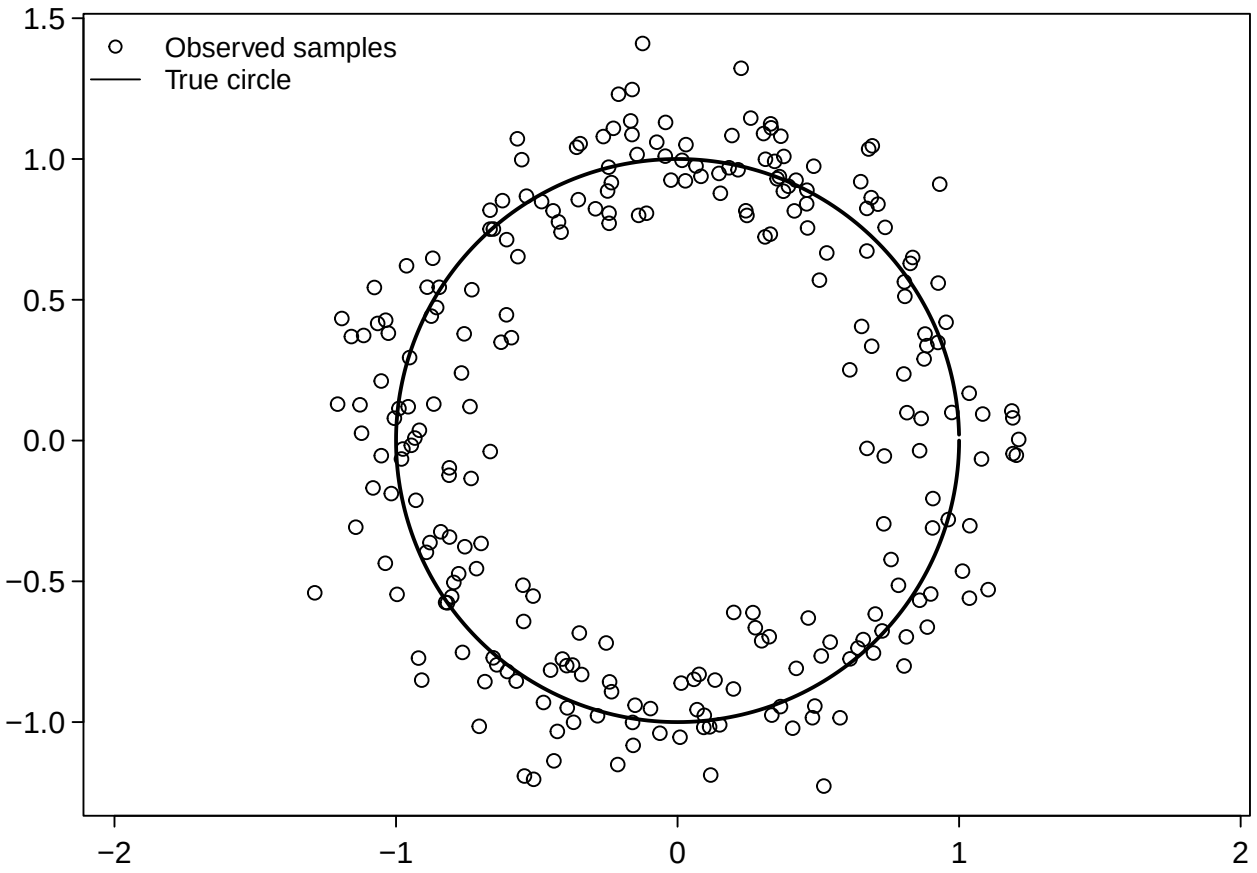
This vignette follows one complete analysis story. We start with noisy samples from a circle and a noisy response field over that circle. We then construct graph scales, denoise geometry, smooth responses and features, recover local extrema/basins, and finally test local associations with permutation-based inference.

1) Simulate Geometry and Response

```
n <- 250L
radius <- 1

X.df <- generate.circle.data(
  n = n,
  radius = radius,
  noise = 0.15,
  type = "random",
  noise.type = "normal",
  seed = 11
)
X <- as.matrix(X.df[, c("x", "y")])

if ("angles" %in% colnames(X.df)) {
  theta.obs <- as.numeric(X.df$angles)
} else {
  theta.obs <- atan2(X[, 2], X[, 1])
  theta.obs <- ifelse(theta.obs < 0, theta.obs + 2 * pi, theta.obs)
}

# Response: a mixture of two circular Gaussian peaks (one larger, one smaller)
mu1 <- 0.80
mu2 <- 3.95
sd1 <- 0.30
sd2 <- 0.48
amp1 <- 1.40
amp2 <- 0.75

comp1 <- circular.synthetic.mixture.of.gaussians(
  x = theta.obs,
  x.knot = mu1,
  y.knot = amp1,
  sd.knot = sd1
)
comp2 <- circular.synthetic.mixture.of.gaussians(
  x = theta.obs,
  x.knot = mu2,
  y.knot = amp2,
```

```
  sd.knot = sd2
)
y.true <- comp1 + comp2
y <- y.true + rnorm(n, sd = 0.10)

circle.true.df <- generate.circle.data(
  n = 300,
  radius = radius,
  noise = 0,
  type = "uniform",
  noise.type = "normal",
  seed = 1
)
circle.true <- as.matrix(circle.true.df[, c("x", "y")])
```

```
op <- par(mar = c(2.5, 2.5, 0.5, 0.5), mgp = c(2.5, 0.5, 0), tcl = -0.3)
plot(X, asp = 1, las = 1, xlab = "", ylab = "")
lines(circle.true, lwd = 2)
legend("topleft",
       legend = c("Observed samples", "True circle"),
       pch = c(1, NA),
       lty = c(NA, 1),
       col = c("black", "black"),
       bty = "n", cex = 0.9)
```



```
par(op)
```

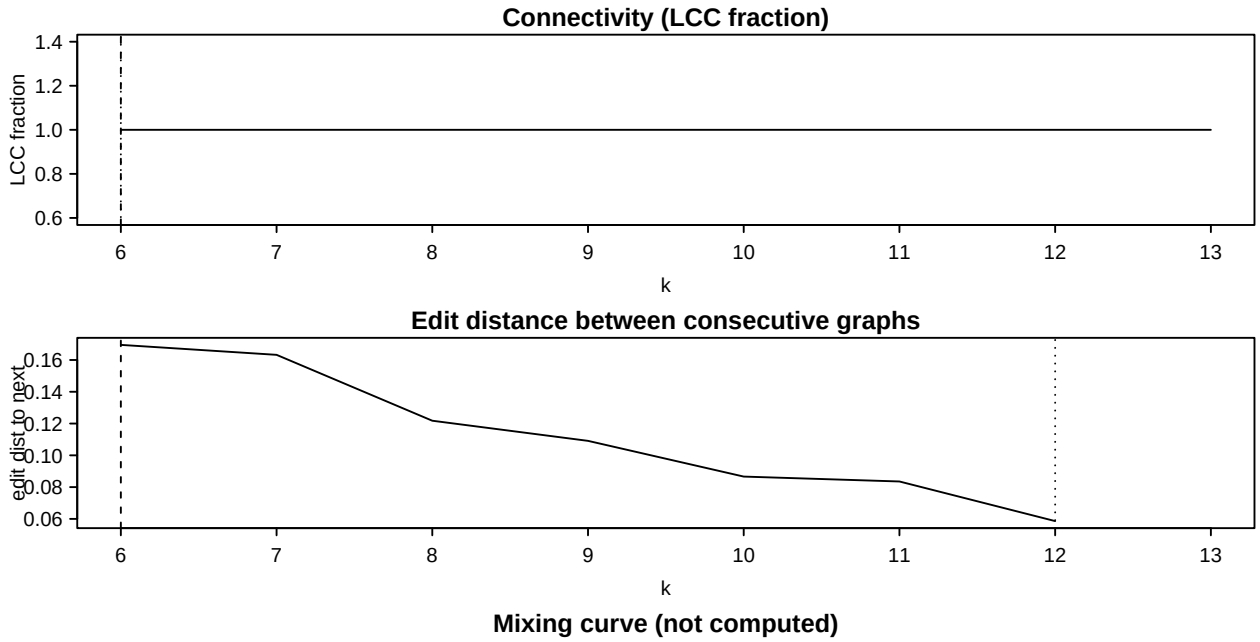
2) Build Graphs Across k

```
k.min <- 5L
k.max <- 12L

X.graphs <- create.iknn.graphs(
  X,
  kmin = k.min,
  kmax = k.max,
  max.path.edge.ratio.deviation.thld = 0.1,
  n.cores = 1L,
  verbose = TRUE
)
#>  Using 1 OpenMP thread
#> Processing k values from 5 to 12 for 250 vertices
#> Requested parallel.mode: auto
#> Parallel mode: serial
#> Starting graph processing
#> [compute_knn] running...[compute_knn] 0.001s
#> [graph_build/geometric_prune/isize_prune] running...[graph_build/geometric_prune/isize_prune] 0.381s
#> Graph processing completed (0.381)
#> [compute_knn] 0.001s
#> [graph_build] 0.014s
#> [geometric_prune] 0.367s
#> [isize_prune] skipped (with.isize.pruning=FALSE)
#> Creating return list objects ... DONE (0.000)
#> Total elapsed time (0.381)
summary(X.graphs)
#> Summary of iknn_graphs object
#> ----------------------------
#> Number of vertices: 250
#> k range: 5 to 12
#> Max path-edge ratio deviation threshold: 0.1
#> Path-edge ratio percentile: 0.5
#> Graph type: geometrically pruned
#>
#>  idx  k n_ccomp edges mean_degree min_degree max_degree sparsity
#>    1  5       1   877        7.02          2         15  0.97182
#>    2  6       1  1056        8.45          3         17  0.96607
#>    3  7       1  1262       10.10          3         22  0.95945
#>    4  8       1  1437       11.50          3         23  0.95383
#>    5  9       1  1613       12.90          3         29  0.94818
#>    6 10       1  1766       14.13          3         29  0.94326
#>    7 11       1  1927       15.42          3         32  0.93809
#>    8 12       1  2047       16.38          3         33  0.93423
```

3) Select Graph Scale

```
X.graphs2 <- build.iknn.graphs.and.selectk(
  X,
```

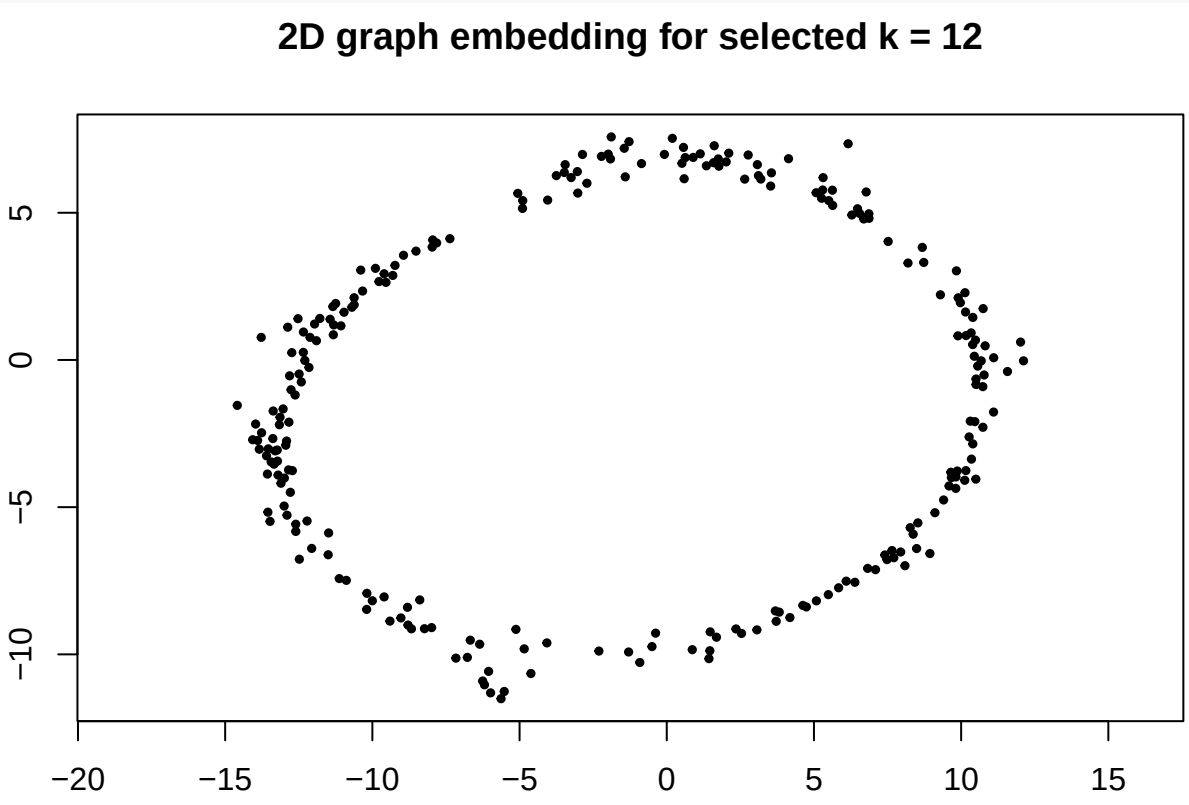
```
  kmin = k.min,
  kmax = k.max,
  method = "edit",
  n.cores = 1L,
  verbose = TRUE
)
#>  Using 1 OpenMP thread
#> Processing k values from 5 to 12 for 250 vertices
#> Requested parallel.mode: auto
#> Parallel mode: serial
#> Starting graph processing
#> [compute_knn] running...[compute_knn] 0.001s
#> [graph_build/geometric_prune/isize_prune] running...[graph_build/geometric_prune/isize_prune] 0.378s
#> Graph processing completed (0.378)
#> [compute_knn] 0.001s
#> [graph_build] 0.013s
#> [geometric_prune] 0.365s
#> [isize_prune] skipped (with.isize.pruning=FALSE)
#> Creating return list objects ... DONE (0.000)
#> Total elapsed time (0.379)

X.graphs2$k.opt.edit
#> [1] 12
plot(X.graphs2)
```



4) Visualize the Selected Graph

```
k.values <- if (!is.null(colnames(X.graphs$k_statistics)) &&
                "k" %in% colnames(X.graphs$k_statistics)) {
  as.integer(X.graphs$k_statistics[, "k"])
} else {
  seq_len(length(X.graphs$geom_pruned_graphs))
```

```
}

sel.k <- X.graphs2$k.opt.edit
if (is.null(sel.k) || is.na(sel.k)) {
  sel.k <- k.values[1]
}
sel.k <- as.integer(sel.k)
k.idx <- which(k.values == sel.k)[1]
if (is.na(k.idx)) k.idx <- 1L

g.sel <- X.graphs$geom_pruned_graphs[[k.idx]]
layout.2d <- graph.embedding(
  adj.list = g.sel$adj_list,
  weights.list = g.sel$weight_list,
  invert.weights = TRUE,
  dim = 2,
  method = "fr"
)

plot(layout.2d, pch = 19, cex = 0.5, asp = 1,
     main = sprintf("2D graph embedding for selected k = %d", sel.k),
     xlab = "", ylab = "")
```



5) Denoise Geometry with `data.smoother()`

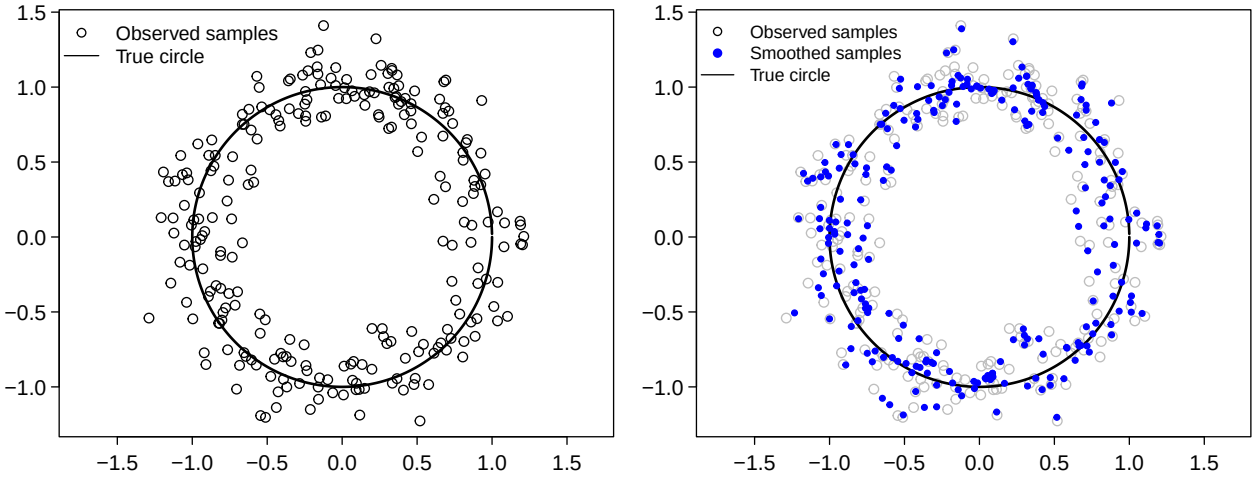
```
X.denoise.res <- data.smoother(
  X,
  kmin = k.min,
```

```
  kmax = k.max,
  n.cores = 1L,
  proxy.response = "pc1",
  max.iterations = 8L,
  n.eigenpairs = 100L,
  filter.type = "heat_kernel",
  t.scale.factor = 0.5,
  beta.coef.factor = 0.1,
  verbose = TRUE
)
#> Building ikNN graphs (geom pruned) for k in [5, 12]
#>  Using 1 OpenMP thread
#> Processing k values from 5 to 12 for 250 vertices
#> Requested parallel.mode: auto
#> Parallel mode: serial
#> Starting graph processing
#> [compute_knn] running...[compute_knn] 0.001s
#> [graph_build/geometric_prune/isize_prune] running...[graph_build/geometric_prune/isize_prune] 0.382s
#> Graph processing completed (0.382)
#> [compute_knn] 0.001s
#> [graph_build] 0.014s
#> [geometric_prune] 0.368s
#> [isize_prune] skipped (with.isize.pruning=FALSE)
#> Creating return list objects ... DONE (0.000)
#> Total elapsed time (0.382)
#> Summary of iknn_graphs object
#> ----------------------------
#> Number of vertices: 250
#> k range: 5 to 12
#> Max path-edge ratio deviation threshold: 0.1
#> Path-edge ratio percentile: 0.5
#> Graph type: geometrically pruned
#>
#>  idx  k n_ccomp edges mean_degree min_degree max_degree sparsity
#>    1  5       1   877        7.02          2         15  0.97182
#>    2  6       1  1056        8.45          3         17  0.96607
#>    3  7       1  1262       10.10          3         22  0.95945
#>    4  8       1  1437       11.50          3         23  0.95383
#>    5  9       1  1613       12.90          3         29  0.94818
#>    6 10       1  1766       14.13          3         29  0.94326
#>    7 11       1  1927       15.42          3         32  0.93809
#>    8 12       1  2047       16.38          3         33  0.93423
#> Proxy response: PC1 score (fraction variance explained = 0.5565)
#> Fitting fit.rdgraph.regression() across k values and extracting GCV
#> Selected k = 5 (min GCV = 1.212e-05)

X.smoothed <- X.denoise.res$X.smoothed
X.denoise.res$k.best
#> [1] 5
```

```
if (isTRUE(X.denoise.res$trimmed)) {
  kept.rows <- X.denoise.res$kept.rows
  X.use <- X[kept.rows, , drop = FALSE]
  theta.use <- theta.obs[kept.rows]
```

```
  y.use <- y[kept.rows]
  y.true.use <- y.true[kept.rows]
  comp1.use <- comp1[kept.rows]
  comp2.use <- comp2[kept.rows]
} else {
  X.use <- X
  theta.use <- theta.obs
  y.use <- y
  y.true.use <- y.true
  comp1.use <- comp1
  comp2.use <- comp2
}
```

```
op <- par(mfrow = c(1, 2),
          mar = c(2.75, 2.75, 0.5, 0.5),
          mgp = c(2.5, 0.5, 0),
          tcl = -0.3)

plot(X.use, las = 1, asp = 1, xlab = "", ylab = "")
lines(circle.true, lwd = 2)
legend("topleft",
       legend = c("Observed samples", "True circle"),
       pch = c(1, NA),
       lty = c(NA, 1),
       col = c("black", "black"),
       bty = "n", cex = 0.9)

plot(X.use, las = 1, asp = 1, col = "gray", xlab = "", ylab = "")
lines(circle.true, lwd = 2)
points(X.smoothed, col = "blue", pch = 19, cex = 0.6)
legend("topleft",
       legend = c("Observed samples", "Smoothed samples", "True circle"),
       pch = c(1, 19, NA),
       lty = c(NA, NA, 1),
       col = c("black", "blue", "black"),
       bty = "n", cex = 0.85)
```




```
par(op)
```

```r
rad.obs <- sqrt(rowSums(X.use^2))
rad.sm <- sqrt(rowSums(X.smoothed^2))

rmse.obs <- sqrt(mean((rad.obs - radius)^2))
rmse.sm <- sqrt(mean((rad.sm - radius)^2))

data.frame(
  metric = c("RMSE radius (observed)", "RMSE radius (smoothed)"),
  value = c(rmse.obs, rmse.sm)
)
#>                    metric     value
#> 1 RMSE radius (observed) 0.1472329
#> 2 RMSE radius (smoothed) 0.1321032
```

6) Smooth the Response on the Selected Graph

We reuse the graph/spectral structure selected by `data.smoother()` and apply it to `y` via `refit.rdgraph.regression()`.

```r
fit.seed <- X.denoise.res$fit.best

y.fit <- refit.rdgraph.regression(
  fitted.model = fit.seed,
  y.new = as.double(y.use),
  per.column.gcv = TRUE,
  n.cores = 1L,
  verbose = FALSE
)

y.hat <- as.double(y.fit$fitted.values)

data.frame(
  metric = c("cor(y.hat, y.true)", "RMSE(y.hat, y.true)"),
  value = c(cor(y.hat, y.true.use),
            sqrt(mean((y.hat - y.true.use)^2)))
)
#>                 metric      value
#> 1  cor(y.hat, y.true) 0.98165081
#> 2 RMSE(y.hat, y.true) 0.07042971
```

Optional 3D Surface View

```r
X.denoise.graph.layout.3d <- graph.embedding(
  adj.list = fit.seed$graph$adj.list,
  weights.list = fit.seed$graph$edge.length.list,
  invert.weights = TRUE,
  dim = 3,
  method = "fr"
)
```

```r
if (requireNamespace("rgl", quietly = TRUE)) {
  plot3D.cont(
    X.denoise.graph.layout.3d,
    y.hat,
    radius = 0.03,
```

```
    legend.title = "y.hat"
  )
}
```

```
if (requireNamespace("rgl", quietly = TRUE) &&
    requireNamespace("htmlwidgets", quietly = TRUE)) {
  plot3D.cont.html(
    X.denoise.graph.layout.3d,
    y.hat,
    legend.title = "y.hat",
    output.file = NULL
  )
}
```

7) Build and Smooth Feature Functions

Features are built so some correlate with selected arcs of the response, while others are pure noise.

```
set.seed(1002)

window.1 <- as.numeric(comp1.use >= quantile(comp1.use, probs = 0.60))
window.2 <- as.numeric(comp2.use >= quantile(comp2.use, probs = 0.60))

z.global <- y.true.use + rnorm(length(theta.use), sd = 0.08)
z.peak1.arc <- window.1 * y.true.use + rnorm(length(theta.use), sd = 0.08)
z.peak2.arc <- window.2 * y.true.use + rnorm(length(theta.use), sd = 0.08)
z.peak.contrast <- (comp1.use - comp2.use) + rnorm(length(theta.use), sd = 0.08)

z.noise1 <- rnorm(length(theta.use))
z.noise2 <- rnorm(length(theta.use))
z.noise3 <- rnorm(length(theta.use))

Z <- cbind(
  z_global = z.global,
  z_peak1_arc = z.peak1.arc,
  z_peak2_arc = z.peak2.arc,
  z_peak_contrast = z.peak.contrast,
  z_noise1 = z.noise1,
  z_noise2 = z.noise2,
  z_noise3 = z.noise3
)
Z <- scale(Z)
```

```
Z.fit <- refit.rdgraph.regression(
  fitted.model = fit.seed,
  y.new = Z,
  per.column.gcv = FALSE,
  n.cores = 1L,
  verbose = FALSE
)
Z.sm <- as.matrix(Z.fit$fitted.values)
summary(Z.fit)
#>
#> Summary: Refitted Riemannian Graph Regression
#> ============================================
```

```
#>
#> Observations: 250
#> Responses: 7
#>
#> Fit quality summary across 7 responses:
#>   R-sq: min=0.3491, Q1=0.3768, med=0.9401, Q3=0.9621, max=0.9775
#>   RMSE: min=0.1499, Q1=0.1941, med=0.2442, Q3=0.7879, max=0.8051
```

8) Recover Extrema and Basins

```
y.basins <- compute.refined.basins(
  adj.list = fit.seed$graph$adj.list,
  edge.length.list = fit.seed$graph$edge.length.list,
  fitted.values = y.hat,
  edge.length.quantile.thld = 0.8,
  min.rel.value.max = 1.05,
  max.rel.value.min = 0.95,
  max.overlap.threshold = 0.20,
  min.overlap.threshold = 0.20,
  p.mean.nbrs.dist.threshold = 0.90,
  p.mean.hopk.dist.threshold = 0.90,
  p.deg.threshold = 0.90,
  min.basin.size = 8L,
  expand.basins = TRUE,
  apply.geometric.filter = TRUE,
  with.trajectories = TRUE,
  verbose = TRUE
)
#> Step 1: Computing initial basins of attraction...
#>   Found 16 maxima and 15 minima
#> initial.summary:
#>    label vertex       value   rel.value type hop.idx basin.size
#> 1     m1     92 -0.07435520 -0.23225749  min       9         51
#> 2     m2    120 -0.07235936 -0.22602322  min      10         33
#> 3     m3    245 -0.06818027 -0.21296932  min       2          4
#> 4     m4     78 -0.05805448 -0.18134021  min       8         47
#> 5     m5    230 -0.05711243 -0.17839760  min       4         17
#> 6     m6     93 -0.04972861 -0.15533335  min       3          6
#> 7     m7    250 -0.01069331 -0.03340185  min       1          6
#> 8     m8    208  0.01596320  0.04986297  min       1          4
#> 9     m9    211  0.05125932  0.16011471  min       1          6
#> 10   m10      6  0.05695806  0.17791540  min       4         13
#> 11   m11     60  0.27918710  0.87207471  min       0          1
#> 12   m12    180  0.49723288  1.55316714  min       5          8
#> 13   m13    168  0.70244102  2.19415962  min       1          6
#> 14   m14     33  1.13802412  3.55475619  min       2          5
#> 15   m15     23  1.22327536  3.82104877  min       0          1
#> 16    M1     30  1.31609441  4.11098029  max       9         48
#> 17    M2     28  1.29445776  4.04339560  max       5         14
#> 18    M3     20  1.25101199  3.90768748  max       3         10
#> 19    M4    163  0.81139899  2.53450303  max       3          4
#> 20    M5    148  0.65358794  2.04156109  max       0          1
#> 21    M6    129  0.19186557  0.59931535  max       5          8
```

```
#> 22    M7      8  0.17613234  0.55017071  max       3          5
#> 23    M8    217  0.14681274  0.45858737  max       1          6
#> 24    M9    103  0.14119784  0.44104856  max       2          7
#> 25   M10     77  0.11528141  0.36009545  max       4         18
#> 26   M11    215  0.11506226  0.35941091  max       1          5
#> 27   M12     71  0.11308360  0.35323032  max       5         11
#> 28   M13    247  0.10453791  0.32653683  max       3          8
#> 29   M14    239  0.05751341  0.17965011  max       4         13
#> 30   M15    233  0.05294986  0.16539529  max       2         10
#> 31   M16    102  0.03321426  0.10374875  max       3          8
#>    p.mean.nbrs.dist p.mean.hopk.dist deg p.deg
#> 1             0.168            0.092   6 0.684
#> 2             0.100            0.512   5 0.896
#> 3             0.788            0.912   4 0.956
#> 4             0.404            0.648   6 0.684
#> 5             0.696            0.364   7 0.504
#> 6             0.812            0.720   5 0.896
#> 7             0.016            0.380   5 0.896
#> 8             0.780            0.964   5 0.896
#> 9             0.732            0.572   6 0.684
#> 10            0.496            0.224   5 0.896
#> 11            0.972            0.676   2 1.000
#> 12            0.664            0.976   3 0.980
#> 13            0.008            0.012   5 0.896
#> 14            0.624            0.952   5 0.896
#> 15            0.996            0.996   4 0.956
#> 16            0.328            0.208   3 0.980
#> 17            0.096            0.616   6 0.684
#> 18            0.564            0.700   6 0.684
#> 19            0.628            0.288   2 1.000
#> 20            1.000            1.000   3 0.980
#> 21            0.244            0.008   2 1.000
#> 22            0.568            0.048   3 0.980
#> 23            0.304            0.388   6 0.684
#> 24            0.548            0.472   5 0.896
#> 25            0.408            0.232   5 0.896
#> 26            0.144            0.576   5 0.896
#> 27            0.988            0.992   2 1.000
#> 28            0.032            0.504   5 0.896
#> 29            0.804            0.552   5 0.896
#> 30            0.240            0.096   5 0.896
#> 31            0.020            0.756   5 0.896
#>
#> Step 2: Filtering by relative values (max >= 1.05, min <= 0.95)...
#>   Retained 5 maxima and 11 minima
#> relvalue.summary:
#>    label vertex       value   rel.value type hop.idx basin.size
#> 1     m1     92 -0.07435520 -0.23225749  min       9         51
#> 2     m2    120 -0.07235936 -0.22602322  min      10         33
#> 3     m3    245 -0.06818027 -0.21296932  min       2          4
#> 4     m4     78 -0.05805448 -0.18134021  min       8         47
#> 5     m5    230 -0.05711243 -0.17839760  min       4         17
#> 6     m6     93 -0.04972861 -0.15533335  min       3          6
```

```
#> 7     m7    250 -0.01069331 -0.03340185  min       1          6
#> 8     m8    208  0.01596320  0.04986297  min       1          4
#> 9     m9    211  0.05125932  0.16011471  min       1          6
#> 10   m10      6  0.05695806  0.17791540  min       4         13
#> 11   m11     60  0.27918710  0.87207471  min       0          1
#> 12    M1     30  1.31609441  4.11098029  max       9         48
#> 13    M2     28  1.29445776  4.04339560  max       5         14
#> 14    M3     20  1.25101199  3.90768748  max       3         10
#> 15    M4    163  0.81139899  2.53450303  max       3          4
#> 16    M5    148  0.65358794  2.04156109  max       0          1
#>    p.mean.nbrs.dist p.mean.hopk.dist deg p.deg
#> 1             0.168            0.092   6 0.684
#> 2             0.100            0.512   5 0.896
#> 3             0.788            0.912   4 0.956
#> 4             0.404            0.648   6 0.684
#> 5             0.696            0.364   7 0.504
#> 6             0.812            0.720   5 0.896
#> 7             0.016            0.380   5 0.896
#> 8             0.780            0.964   5 0.896
#> 9             0.732            0.572   6 0.684
#> 10            0.496            0.224   5 0.896
#> 11            0.972            0.676   2 1.000
#> 12            0.328            0.208   3 0.980
#> 13            0.096            0.616   6 0.684
#> 14            0.564            0.700   6 0.684
#> 15            0.628            0.288   2 1.000
#> 16            1.000            1.000   3 0.980
#>
#> Step 3: Clustering maxima (overlap threshold = 0.20)...
#>   Found 5 clusters (5 singletons, 0 to be merged)
#>   Merging clustered maxima...
#>   Result: 5 maxima after merging
#> merged.max.summary:
#>    label vertex       value   rel.value type hop.idx basin.size
#> 1     m1     92 -0.07435520 -0.23225749  min       9         51
#> 2     m2    120 -0.07235936 -0.22602322  min      10         33
#> 3     m3    245 -0.06818027 -0.21296932  min       2          4
#> 4     m4     78 -0.05805448 -0.18134021  min       8         47
#> 5     m5    230 -0.05711243 -0.17839760  min       4         17
#> 6     m6     93 -0.04972861 -0.15533335  min       3          6
#> 7     m7    250 -0.01069331 -0.03340185  min       1          6
#> 8     m8    208  0.01596320  0.04986297  min       1          4
#> 9     m9    211  0.05125932  0.16011471  min       1          6
#> 10   m10      6  0.05695806  0.17791540  min       4         13
#> 11   m11     60  0.27918710  0.87207471  min       0          1
#> 12    M1     30  1.31609441  4.11098029  max       9         48
#> 13    M2     28  1.29445776  4.04339560  max       5         14
#> 14    M3     20  1.25101199  3.90768748  max       3         10
#> 15    M4    163  0.81139899  2.53450303  max       3          4
#> 16    M5    148  0.65358794  2.04156109  max       0          1
#>    p.mean.nbrs.dist p.mean.hopk.dist deg p.deg
#> 1             0.168            0.092   6 0.684
#> 2             0.100            0.512   5 0.896
```

```
#> 3             0.788            0.912   4 0.956
#> 4             0.404            0.648   6 0.684
#> 5             0.696            0.364   7 0.504
#> 6             0.812            0.720   5 0.896
#> 7             0.016            0.380   5 0.896
#> 8             0.780            0.964   5 0.896
#> 9             0.732            0.572   6 0.684
#> 10            0.496            0.224   5 0.896
#> 11            0.972            0.676   2 1.000
#> 12            0.328            0.208   3 0.980
#> 13            0.096            0.616   6 0.684
#> 14            0.564            0.700   6 0.684
#> 15            0.628            0.288   2 1.000
#> 16            1.000            1.000   3 0.980
#>
#> Step 4: Clustering minima (overlap threshold = 0.20)...
#>   Found 10 clusters (9 singletons, 1 to be merged)
#>   Merging clustered minima...
#>   Result: 10 minima after merging
#> merged.min.summary:
#>    label vertex       value   rel.value type hop.idx basin.size
#> 1     m1     92 -0.07435520 -0.23225749  min       9         58
#> 2     m2    120 -0.07235936 -0.22602322  min      10         33
#> 3     m3    245 -0.06818027 -0.21296932  min       2          4
#> 4     m4    230 -0.05711243 -0.17839760  min       4         17
#> 5     m5     93 -0.04972861 -0.15533335  min       3          6
#> 6     m6    250 -0.01069331 -0.03340185  min       1          6
#> 7     m7    208  0.01596320  0.04986297  min       1          4
#> 8     m8    211  0.05125932  0.16011471  min       1          6
#> 9     m9      6  0.05695806  0.17791540  min       4         13
#> 10   m10     60  0.27918710  0.87207471  min       0          1
#> 11    M1     30  1.31609441  4.11098029  max       9         48
#> 12    M2     28  1.29445776  4.04339560  max       5         14
#> 13    M3     20  1.25101199  3.90768748  max       3         10
#> 14    M4    163  0.81139899  2.53450303  max       3          4
#> 15    M5    148  0.65358794  2.04156109  max       0          1
#>    p.mean.nbrs.dist p.mean.hopk.dist deg p.deg
#> 1             0.168            0.092   6 0.684
#> 2             0.100            0.512   5 0.896
#> 3             0.788            0.912   4 0.956
#> 4             0.696            0.364   7 0.504
#> 5             0.812            0.720   5 0.896
#> 6             0.016            0.380   5 0.896
#> 7             0.780            0.964   5 0.896
#> 8             0.732            0.572   6 0.684
#> 9             0.496            0.224   5 0.896
#> 10            0.972            0.676   2 1.000
#> 11            0.328            0.208   3 0.980
#> 12            0.096            0.616   6 0.684
#> 13            0.564            0.700   6 0.684
#> 14            0.628            0.288   2 1.000
#> 15            1.000            1.000   3 0.980
#>
```

```
#> Step 5: Filtering by geometric characteristics and basin size:
#>   p.mean.nbrs.dist < 0.90, p.mean.hopk.dist < 0.90, p.deg < 0.90, basin.size >= 8 ...
#>   Maxima: 5 -> 2 (removed 3)
#> geom.summary:
#>    label vertex       value   rel.value type hop.idx basin.size
#> 1     m1     92 -0.07435520 -0.23225749  min       9         58
#> 2     m2    120 -0.07235936 -0.22602322  min      10         33
#> 3     m3    245 -0.06818027 -0.21296932  min       2          4
#> 4     m4    230 -0.05711243 -0.17839760  min       4         17
#> 5     m5     93 -0.04972861 -0.15533335  min       3          6
#> 6     m6    250 -0.01069331 -0.03340185  min       1          6
#> 7     m7    208  0.01596320  0.04986297  min       1          4
#> 8     m8    211  0.05125932  0.16011471  min       1          6
#> 9     m9      6  0.05695806  0.17791540  min       4         13
#> 10   m10     60  0.27918710  0.87207471  min       0          1
#> 11    M1     30  1.31609441  4.11098029  max       9         48
#> 12    M2     28  1.29445776  4.04339560  max       5         14
#> 13    M3     20  1.25101199  3.90768748  max       3         10
#> 14    M4    163  0.81139899  2.53450303  max       3          4
#> 15    M5    148  0.65358794  2.04156109  max       0          1
#>    p.mean.nbrs.dist p.mean.hopk.dist deg p.deg
#> 1             0.168            0.092   6 0.684
#> 2             0.100            0.512   5 0.896
#> 3             0.788            0.912   4 0.956
#> 4             0.696            0.364   7 0.504
#> 5             0.812            0.720   5 0.896
#> 6             0.016            0.380   5 0.896
#> 7             0.780            0.964   5 0.896
#> 8             0.732            0.572   6 0.684
#> 9             0.496            0.224   5 0.896
#> 10            0.972            0.676   2 1.000
#> 11            0.328            0.208   3 0.980
#> 12            0.096            0.616   6 0.684
#> 13            0.564            0.700   6 0.684
#> 14            0.628            0.288   2 1.000
#> 15            1.000            1.000   3 0.980
#>   Minima: 10 -> 4 (removed 6)
#>
#> Generating final summary (hop.k = 2)...
#> Populating extrema labels in basins object...
#> final.summary:
#>   label vertex       value  rel.value type hop.idx basin.size p.mean.nbrs.dist
#> 1    m1     92 -0.07435520 -0.2322575  min       9         58            0.168
#> 2    m2    120 -0.07235936 -0.2260232  min      10         33            0.100
#> 3    m3    230 -0.05711243 -0.1783976  min       4         17            0.696
#> 4    m4      6  0.05695806  0.1779154  min       4         13            0.496
#> 5    M1     28  1.29445776  4.0433956  max       5         14            0.096
#> 6    M2     20  1.25101199  3.9076875  max       3         10            0.564
#>   p.mean.hopk.dist deg p.deg
#> 1            0.092   6 0.684
#> 2            0.512   5 0.896
#> 3            0.364   7 0.504
#> 4            0.224   5 0.896
```

```
#> 5            0.616   6 0.684
#> 6            0.700   6 0.684
#>   Assigned 2 maximum labels
#>   Assigned 4 minimum labels
#> Storing graph structure in basins object...
#> Constructing basin vertices lists...
#> Computing final overlap distance matrices...
#>
#> Step 6: Expanding basins to cover all vertices...
#>   Maxima basins: 17 -> 250 vertices covered
#>   Minima basins: 121 -> 250 vertices covered
#>
#> Refinement complete!
#> Final structure: 2 maxima and 4 minima

y.basins$summary
#>   label vertex       value  rel.value type hop.idx basin.size p.mean.nbrs.dist
#> 1    m1     92 -0.07435520 -0.2322575  min       9         58            0.168
#> 2    m2    120 -0.07235936 -0.2260232  min      10         33            0.100
#> 3    m3    230 -0.05711243 -0.1783976  min       4         17            0.696
#> 4    m4      6  0.05695806  0.1779154  min       4         13            0.496
#> 5    M1     28  1.29445776  4.0433956  max       5         14            0.096
#> 6    M2     20  1.25101199  3.9076875  max       3         10            0.564
#>   p.mean.hopk.dist deg p.deg
#> 1            0.092   6 0.684
#> 2            0.512   5 0.896
#> 3            0.364   7 0.504
#> 4            0.224   5 0.896
#> 5            0.616   6 0.684
#> 6            0.700   6 0.684
table(y.basins$summary$type)
#>
#> max min
#>   2   4
```

```
true.max.label <- ifelse(comp1.use >= comp2.use, "M1.true", "M2.true")

max.vertices.list <- y.basins$expanded.max.vertices.list
if (is.null(max.vertices.list)) max.vertices.list <- y.basins$max.vertices.list

est.max.label <- rep(NA_character_, length(y.hat))
for (lab in names(max.vertices.list)) {
  est.max.label[as.integer(max.vertices.list[[lab]])] <- lab
}

idx.ok <- !is.na(est.max.label)
table(True = true.max.label[idx.ok], Estimated = est.max.label[idx.ok])
#>          Estimated
#> True       M1 M2
#>   M1.true 67 34
#>   M2.true 71 78
```

9) Basin-Wise Pearson and Spearman Checks

```
max.vertices.list <- y.basins$expanded.max.vertices.list
if (is.null(max.vertices.list)) max.vertices.list <- y.basins$max.vertices.list
min.vertices.list <- y.basins$expanded.min.vertices.list
if (is.null(min.vertices.list)) min.vertices.list <- y.basins$min.vertices.list

max.assign <- rep(NA_character_, length(y.hat))
for (lab in names(max.vertices.list)) {
  max.assign[as.integer(max.vertices.list[[lab]])] <- lab
}

min.assign <- rep(NA_character_, length(y.hat))
for (lab in names(min.vertices.list)) {
  min.assign[as.integer(min.vertices.list[[lab]])] <- lab
}

cell.assign <- ifelse(is.na(max.assign) | is.na(min.assign),
                      NA_character_,
                      paste(max.assign, min.assign, sep = "|"))

group.levels <- setdiff(sort(unique(max.assign)), NA_character_)
basin.cor <- data.frame()

for (j in seq_len(ncol(Z.sm))) {
  for (gr in group.levels) {
    idx <- which(max.assign == gr)
    if (length(idx) < 20L) next
    basin.cor <- rbind(
      basin.cor,
      data.frame(
        feature = colnames(Z.sm)[j],
        group = gr,
        n = length(idx),
        pearson = suppressWarnings(cor(y.hat[idx], Z.sm[idx, j], method = "pearson")),
        spearman = suppressWarnings(cor(y.hat[idx], Z.sm[idx, j], method = "spearman"))
      )
    )
  }
}

if (nrow(basin.cor) > 0) {
  head(basin.cor[order(-abs(basin.cor$spearman)), ], 12)
} else {
  basin.cor
}
#>            feature group   n      pearson    spearman
#> 1         z_global    M1 138  0.984200127  0.92080623
#> 2         z_global    M2 112  0.978993761  0.90795093
#> 6      z_peak2_arc    M2 112  0.542105760  0.54193289
#> 3      z_peak1_arc    M1 138  0.841045129  0.44076678
#> 4      z_peak1_arc    M2 112  0.657587352  0.43298767
#> 8  z_peak_contrast    M2 112  0.052568261 -0.30689855
#> 5      z_peak2_arc    M1 138  0.226000979  0.22551936
```

```
#> 7  z_peak_contrast    M1 138  0.588600541  0.19957807
#> 11        z_noise2    M1 138  0.214840600  0.18710537
#> 13        z_noise3    M1 138 -0.117701831 -0.14039850
#> 10        z_noise1    M2 112 -0.002243797  0.08416476
#> 14        z_noise3    M2 112 -0.003026383  0.08313971
```

10) Local Correlation with `lcor()`

```
adj.list <- fit.seed$graph$adj.list
edge.length.list <- fit.seed$graph$edge.length.list

lcor.derivative <- as.matrix(lcor(adj.list, edge.length.list, y.hat, Z, type = "derivative"))
lcor.unit <- as.matrix(lcor(adj.list, edge.length.list, y.hat, Z, type = "unit"))
lcor.sign <- as.matrix(lcor(adj.list, edge.length.list, y.hat, Z, type = "sign"))

lcor.summary <- rbind(
  data.frame(
    feature = colnames(lcor.derivative),
    type = "derivative",
    mean.lcor = colMeans(lcor.derivative),
    mean.abs.lcor = colMeans(abs(lcor.derivative)),
    row.names = NULL
  ),
  data.frame(
    feature = colnames(lcor.unit),
    type = "unit",
    mean.lcor = colMeans(lcor.unit),
    mean.abs.lcor = colMeans(abs(lcor.unit)),
    row.names = NULL
  ),
  data.frame(
    feature = colnames(lcor.sign),
    type = "sign",
    mean.lcor = colMeans(lcor.sign),
    mean.abs.lcor = colMeans(abs(lcor.sign)),
    row.names = NULL
  )
)

lcor.summary[order(lcor.summary$type, -lcor.summary$mean.abs.lcor), ]
#>            feature       type    mean.lcor mean.abs.lcor
#> 1         z_global derivative  0.235500924     0.5245986
#> 2      z_peak1_arc derivative  0.178356229     0.5072634
#> 4  z_peak_contrast derivative -0.007643783     0.4950112
#> 3      z_peak2_arc derivative  0.188635454     0.4824649
#> 6        z_noise2 derivative -0.027890222     0.4539506
#> 7        z_noise3 derivative -0.020302207     0.4460997
#> 5        z_noise1 derivative  0.006122076     0.4445198
#> 15        z_global       sign  0.297452406     0.4843198
#> 18 z_peak_contrast       sign  0.009259557     0.4819137
#> 16     z_peak1_arc       sign  0.203882895     0.4545019
#> 17     z_peak2_arc       sign  0.165927797     0.4066800
#> 21       z_noise3       sign -0.033499020     0.3768995
```

```
#> 19        z_noise1       sign  0.040617490     0.3687459
#> 20        z_noise2       sign -0.031020150     0.3579065
#> 8         z_global       unit  0.297452406     0.4843198
#> 11 z_peak_contrast       unit  0.009259557     0.4819137
#> 9      z_peak1_arc       unit  0.203882895     0.4545019
#> 10     z_peak2_arc       unit  0.165927797     0.4066800
#> 14        z_noise3       unit -0.033499020     0.3768995
#> 12        z_noise1       unit  0.040617490     0.3687459
#> 13        z_noise2       unit -0.031020150     0.3579065
```

11) Permutation Inference for `lcor`

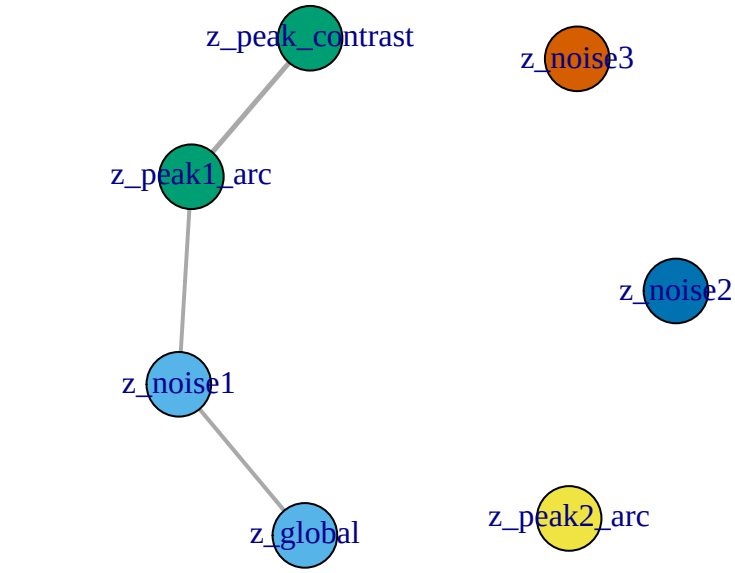
We assess feature significance by permuting feature rows (`permute = "z"`), recomputing local correlations, and comparing observed statistics to this null.

```
perm.lcor <- permutation.test.lcor(
  adj.list = adj.list,
  weight.list = edge.length.list,
  y = y.hat,
  z = Z,
  type = "derivative",
  statistic = "mean.abs",
  permute = "z",
  n.perm = 200L,
  seed = 2026L
)

perm.lcor$table[order(perm.lcor$table$q.value), ]
#>           feature  stat.obs     p.value    q.value
#> 1        z_global 0.5245986 0.004975124 0.01160862
#> 2     z_peak1_arc 0.5072634 0.004975124 0.01160862
#> 4 z_peak_contrast 0.4950112 0.004975124 0.01160862
#> 3     z_peak2_arc 0.4824649 0.014925373 0.02611940
#> 5        z_noise1 0.4445198 0.656716418 0.69154229
#> 6        z_noise2 0.4539506 0.527363184 0.69154229
#> 7        z_noise3 0.4460997 0.691542289 0.69154229
```

12) Build an lcor-Based Feature Module Graph

```
sim <- stats::cor(lcor.derivative, use = "pairwise.complete.obs")
adj.sim <- abs(sim)
diag(adj.sim) <- 0

threshold <- 0.20
adj.sim[adj.sim < threshold] <- 0

if (sum(adj.sim > 0) == 0) {
  message("No edges above threshold. Lower `threshold` to make the module graph denser.")
} else {
  g.mod <- igraph::graph_from_adjacency_matrix(
    adj.sim,
    mode = "undirected",
    weighted = TRUE,
```

```
    diag = FALSE
  )

  modules <- igraph::cluster_louvain(g.mod, weights = igraph::E(g.mod)$weight)

  igraph::plot.igraph(
    g.mod,
    vertex.label = igraph::V(g.mod)$name,
    vertex.size = 26,
    vertex.color = as.integer(igraph::membership(modules)) + 1,
    edge.width = 1 + 4 * igraph::E(g.mod)$weight,
    main = "Feature modules from local-correlation profiles"
  )

  data.frame(
    feature = igraph::V(g.mod)$name,
    module = as.integer(igraph::membership(modules)),
    row.names = NULL
  )
}
```




```
#>           feature module
#> 1        z_global      1
#> 2     z_peak1_arc      2
#> 3     z_peak2_arc      3
#> 4 z_peak_contrast      2
#> 5        z_noise1      1
#> 6        z_noise2      4
#> 7        z_noise3      5
```

Optional `grip` 3D Layout

```
if (requireNamespace("grip", quietly = TRUE) &&
    requireNamespace("rgl", quietly = TRUE)) {
  g <- X.graphs$geom_pruned_graphs[[k.idx]]

  X.graphs.3d <- grip::grip.layout(
    adj_list = g$adj_list,
    weight_list = g$weight_list,
    dim = 3,
    rounds = 50,
    final_rounds = 50,
    num_init = 36,
    num_nbrs = 8,
    seed = 6
  )

  plot3D.plain(X.graphs.3d, radius = 0.03)
}
```

Notes

- For publication-quality inference, increase permutation count (`n.perm`), report confidence intervals, and retain `q.value` (FDR-controlled) summaries.
- For larger datasets, use `n.cores > 1` and consider chunk-level caching.
- If basin refinement is unstable for a specific dataset, start with relaxed filtering and tighten thresholds gradually.