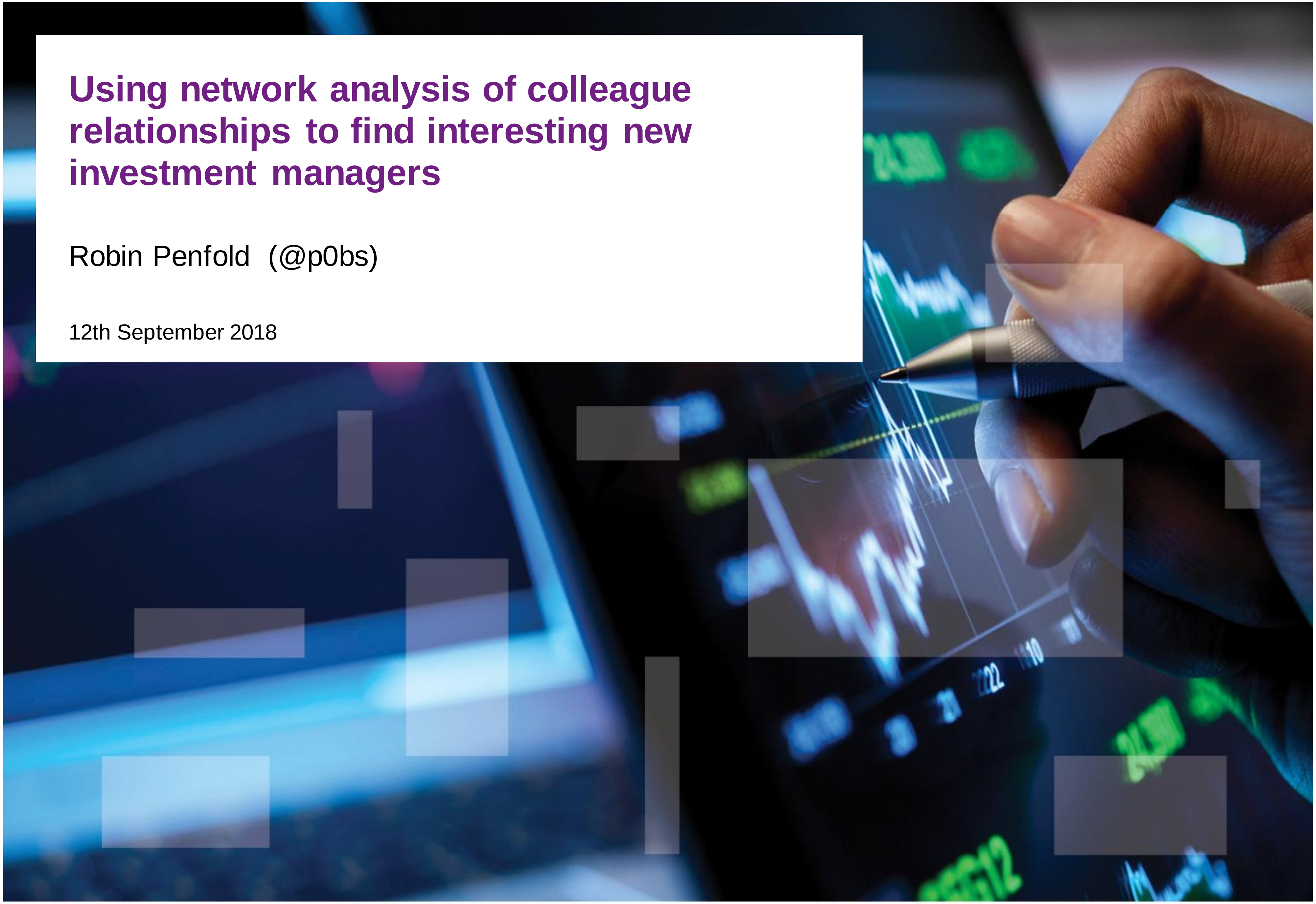
Using network analysis of colleague relationships to find interesting new investment managers

Robin Penfold (@p0bs)

12th September 2018



EARL CONFERENCE

Willis Towers Watson

Why study colleague relationships in investment managers?

- We provide investment services to large institutions, like pension funds
- For example, we recommend which investment managers our clients should use to invest their assets
  - We seek the minority of investment managers that are likely to do better than the market over time
- Our research team therefore interviews many investment managers. But how do we know who to interview?
- One way is to consider those individuals who had an excellent tutor at a previous firm

Bring on the data!

- Industry database of investment managers, showing where they have worked and when
- Dates and current firm are from a drop-down list; previous firms are in free text!
  - “ABC DEF Investment Mgmt, Hong Kong” … not “ABC DEF”
  - “XYZ private equity asset partners in Miami” … not “XYZ”

A computer once beat me at chess
but it was no match for me at kick-boxing

EARL CONFERENCE

Our clean data

- After much de-duplication, we had four columns:
  - newFirmID
  - individual_id
  - start_year
  - end_year

- We can then find the colleague pairings that worked at the same firm in the same year:

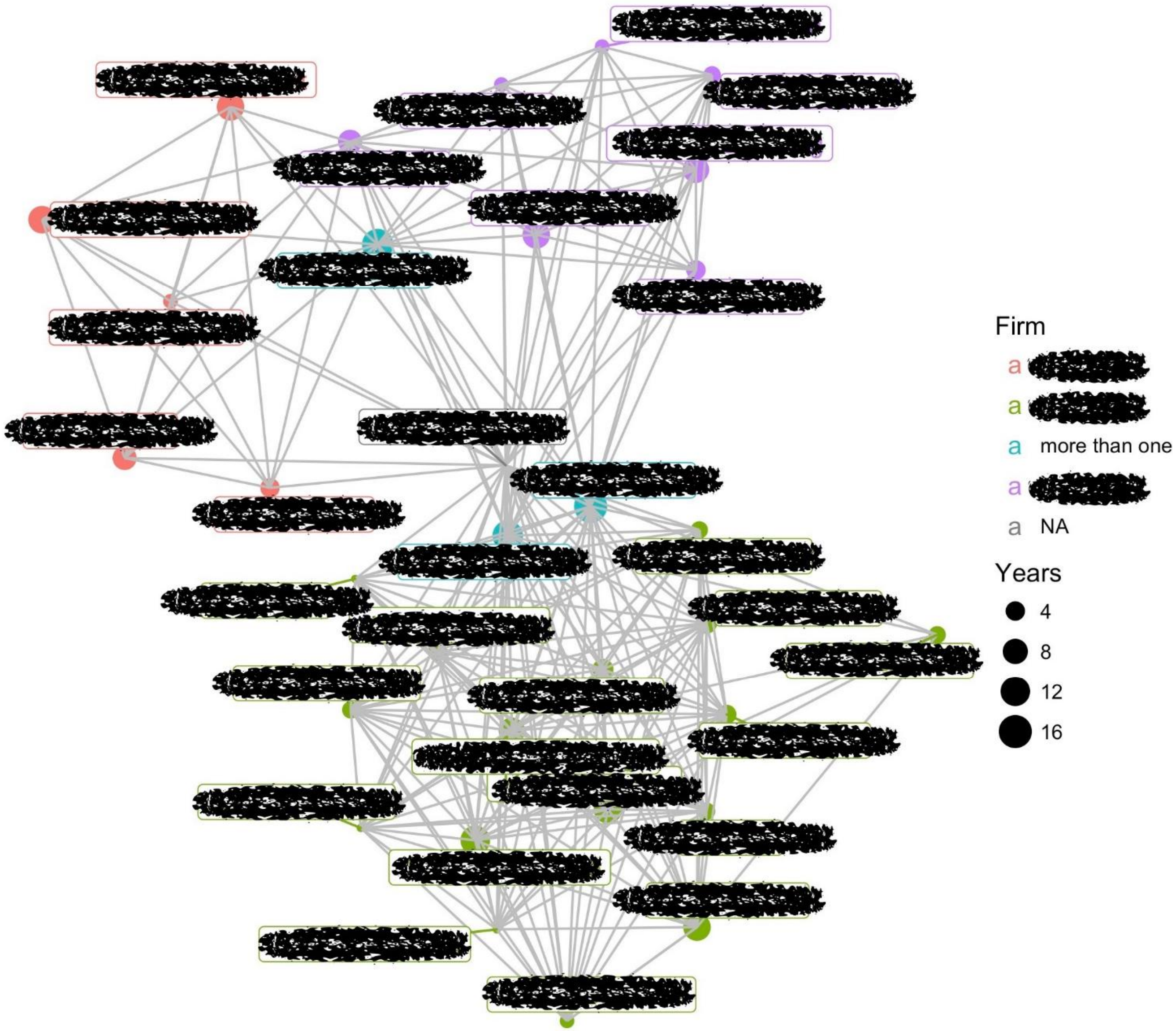
```
> glimpse(colleagues)
Observations: 592,122
Variables: 7
$ newFirmID <int> 9010724, 9009549, 9009549, 9009549, 9009549, 9009549, 9009549, 9009549, 9009549, 9013586, 9013586, 90 ...
$ V1        <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, ...
$ V2        <int> 386, 82497, 2223, 56608, 54621, 1405, 40317, 82463, 47783, 42924, 42926, 1999, 2001, 31619, 6268, 627 ...
$ years     <int> 2, 3, 2, 7, 8, 9, 9, 9, 4, 1, 1, 19, 18, 17, 15, 13, 13, 13, 11, 10, 10, 10, 9, 8, 7, 5, 4, 4, 8, 2,  ...
$ lastYr    <dbl> 1988, 1990, 1991, 1994, 1995, 1996, 1996, 1996, 1996, 1996, 1996, 2014, 2014, 2014, 2014, 2014, 2014, ...
$ FirmLite  <chr> "painewebber", "mitchell hutchins", "mitchell hutchins", "mitchell hutchins", "mitchell hutchins", "m ...
$ firmFREX  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N ...
```

Building the network with tidygraph

```
# A tbl_graph: 12415 nodes and 592122 edges
#
# An undirected multigraph with 26 components
#
# Node Data: 12,415 x 1 (active)
  name
  <chr>
1 2
2 3
3 5
4 6
5 7
6 8
# ... with 1.241e+04 more rows
#
# Edge Data: 592,122 x 6
   from    to firmFREX years lastYr newFirmID
  <int> <int>    <dbl> <int>  <dbl>     <int>
1     1   166       NA     2   1988   9010724
2     1 10386       NA     3   1990   9009549
3     1   707       NA     2   1991   9009549
# ... with 5.921e+05 more rows
```

(Thank you, Thomas Lin Pedersen)

Current and former equity investment colleagues of [redacted]



Which individuals that we don't know have the best former colleagues?

Method 1: Database-like approach

- Find our average 'recommendation rating' across each individual's colleagues

```
tg5 <- tg4 %>%
  activate(nodes) %>%
  mutate(name = as.integer(name)) %>%
  mutate(id = row_number()) %>%
  left_join(EData11_latest, by = c("name" = "InvProID")) %>%
  rename(FirmIDlast = "newFirmID") %>%
  mutate(FirmIDlastFREX = ifelse(
    FirmIDlast %in% InputData1$newFirmID,
    InputData1$firmFREX[match(FirmIDlast, InputData1$newFirmID)],
    NA)) %>%
  mutate(FirmIDlastFREXgd = ifelse(FirmIDlastFREX < goodFREXpoint, FirmIDlast, NA)) %>%
  mutate(neighbours = local_members(order = 1)) %>%
  mutate(colleagueFREXav = map_local_dbl(
    order = 1,
    .f = function(neighborhood, ...) {
      mean(as_tibble(neighborhood, active = 'nodes')$FirmIDlastFREX, na.rm = TRUE)
    })) %>%
  mutate(colleagueFREXgd_distinct = map_local_int(
    order = 1,
    .f = function(neighborhood, ...) {
      length(unique(as_tibble(neighborhood, active = 'nodes')$FirmIDlastFREXgd, na.rm = TRUE))
    }))
```

- Generates a 'best place to look' table of firms and individuals that we don’t yet know

EARL CONFERENCE

Which individuals that we don't know have the best former colleagues?

Method 2: Pure network approach

- Find individuals that we don't know who are most ‘central’ between those we recommend

```
second_network <- tg5 %>%
  activate(nodes) %>%
  mutate(topFirm = FirmIDlastFREX <= topFirmBreakpoint) %>%
  filter(topFirm == TRUE | is.na(topFirm)) %>%
  mutate(wt_node = ifelse(is.na(topFirm), 1, topFirmWeight)) %>%
  activate(edges) %>%
  select(-goodFREXfrom, -goodFREXto, -FREX_average_from, -FREX_average_to) %>%
  mutate(wt_edge = .N()$wt_node[from] + .N()$wt_node[to]) %>%
  activate(nodes) %>%
  mutate(centr_score = centrality_betweenness(weights = wt_edge, directed = FALSE))
```

- Generates a 'best place to look' dataframe of firms and individuals

- Bringing all this together, we can prioritize who to meet first (and at which firms)

Limitations of reliance

Willis Towers Watson has prepared this material for general information purposes only and it should not be considered a substitute for specific professional advice. In particular, its contents are not intended by Willis Towers Watson to be construed as the provision of investment, legal, accounting, tax or other professional advice or recommendations of any kind, or to form the basis of any decision to do or to refrain from doing anything. As such, this material should not be relied upon for investment or other financial decisions and no such decisions should be taken on the basis of its contents without seeking specific advice.

This material is based on information available to Willis Towers Watson at the date of this material and takes no account of subsequent developments after that date. In preparing this material we have relied upon data supplied to us by third parties. Whilst reasonable care has been taken to gauge the reliability of this data, we provide no guarantee as to the accuracy or completeness of this data and Willis Towers Watson and its affiliates and their respective directors, officers and employees accept no responsibility and will not be liable for any errors or misrepresentations in the data made by any third party.

This material may not be reproduced or distributed to any other party, whether in whole or in part, without Willis Towers Watson's prior written permission, except as may be required by law. In the absence of our express written agreement to the contrary, Willis Towers Watson and its affiliates and their respective directors, officers and employees accept no responsibility and will not be liable for any consequences howsoever arising from any use of or reliance on this material or the opinions we have expressed.