

Baseline analysis

Effect of grant size

```
no_network <- df %>%
  filter(network == "none")

no_network_unif_dist <- no_network %>%
  filter(init_dist == "uniform")

pdata <- no_network_unif_dist %>%
  group_by(step, funded_share) %>%
  summarise(mean_gini = mean(resources_gini),
            mean_sharing = mean(perc_sharing))

p1 <- pdata %>%
  ggplot(aes(step, mean_gini, colour = as.factor(funded_share))) +
  geom_line() +
  labs(colour = "% of groups receiving funding",
       y = "Gini of resources")

p2 <- pdata %>%
  ggplot(aes(step, mean_sharing, colour = as.factor(funded_share))) +
  geom_line() +
  labs(colour = "% of groups receiving funding",
       y = "% of groups sharing data")

p1 / p2 +
  plot_layout(guides = "collect") & theme(legend.position = "top")
```

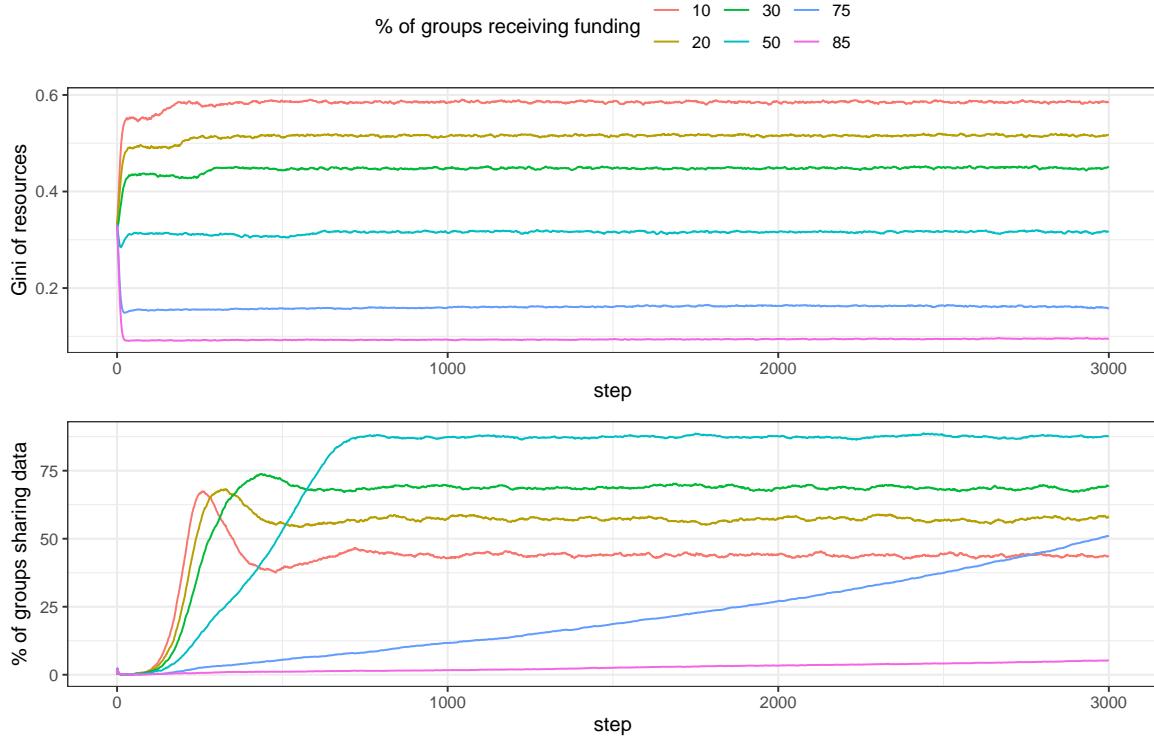


Figure 1: Gini index and % of groups sharing data dependnt on grant size

The above is very interesting: we are not changing incentives, however sharing rate still varies widely. This is a consequence of how exposed agents are to the funding agency. If only few are funded, not many come into contact. However, if almost everyone is funded, the policy seems to work only very slowly, because there is no advantage in sharing or not (because anyways almost everyone is funded). It is also interesting that sharing initially rises, but then drops again (for low values of funded share).

The Gini is in some sense a direct effect of selectivity of funding and thus not particularly interesting when doing this baseline aspect.

Figure 2 visualises variability in the runs.

```
no_network %>%
  # filter(funded_share == 50) %>%
  ggplot(aes(step, perc_sharing, group = run_number,
             colour = as.factor(funded_share))) +
  geom_line(alpha = .2) +
  theme(legend.position = "top") +
```

```

  labs(colour = "% of groups receiving funding",
       y = "% of groups sharing data") +
  guides(colour = guide_legend(override.aes = list(alpha = 1)))

```

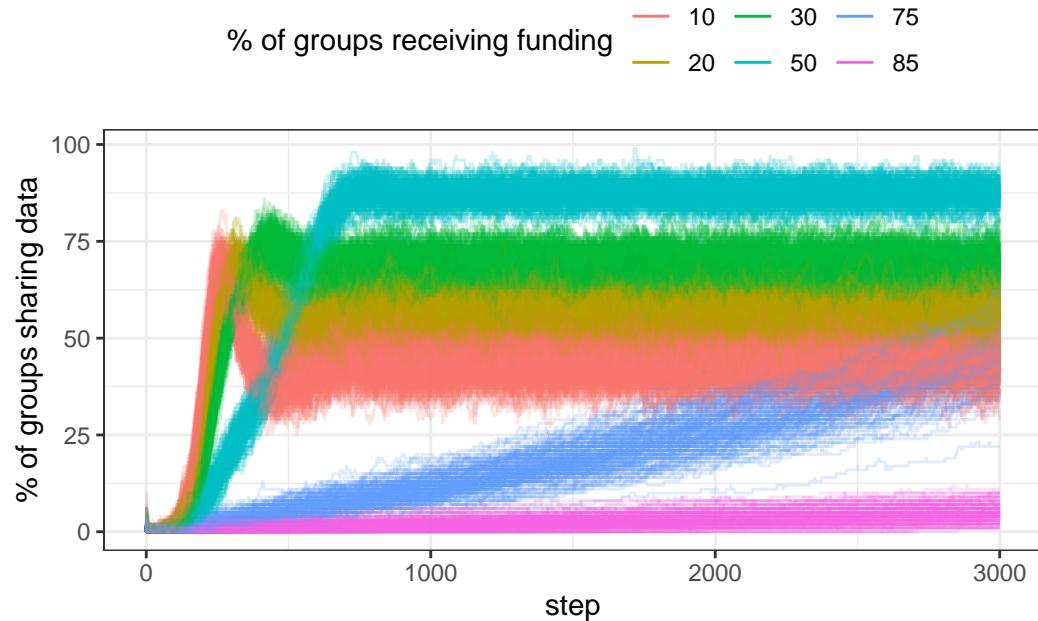


Figure 2: Variability in % of groups sharing data with no network

Comparing network effects

```

uniform <- df %>%
  filter(init_dist == "uniform")

pdata <- uniform %>%
  select(run_number, network, funded_share, step, perc_sharing, resources_gini) %>%
  group_by(network, funded_share, step) %>%
  summarise(mean_gini = mean(resources_gini),
            mean_sharing = mean(perc_sharing)) %>%
  pivot_longer(c(mean_gini, mean_sharing))

p_gini <- pdata %>%
  filter(name == "mean_gini") %>%

```

```
ggplot(aes(step, value, colour = network)) +
  geom_line() +
  facet_wrap(vars(funded_share), ncol = 1) +
  labs(y = "Mean Gini")

p_sharing <- pdata %>%
  filter(name == "mean_sharing") %>%
  ggplot(aes(step, value, colour = network)) +
  geom_line() +
  facet_wrap(vars(funded_share), ncol = 1) +
  labs(y = "Mean % of teams sharing data")

p_sharing + p_gini +
  plot_layout(guides = "collect") +
  plot_annotation(tag_levels = "A") &
  theme(legend.position = "top")
```

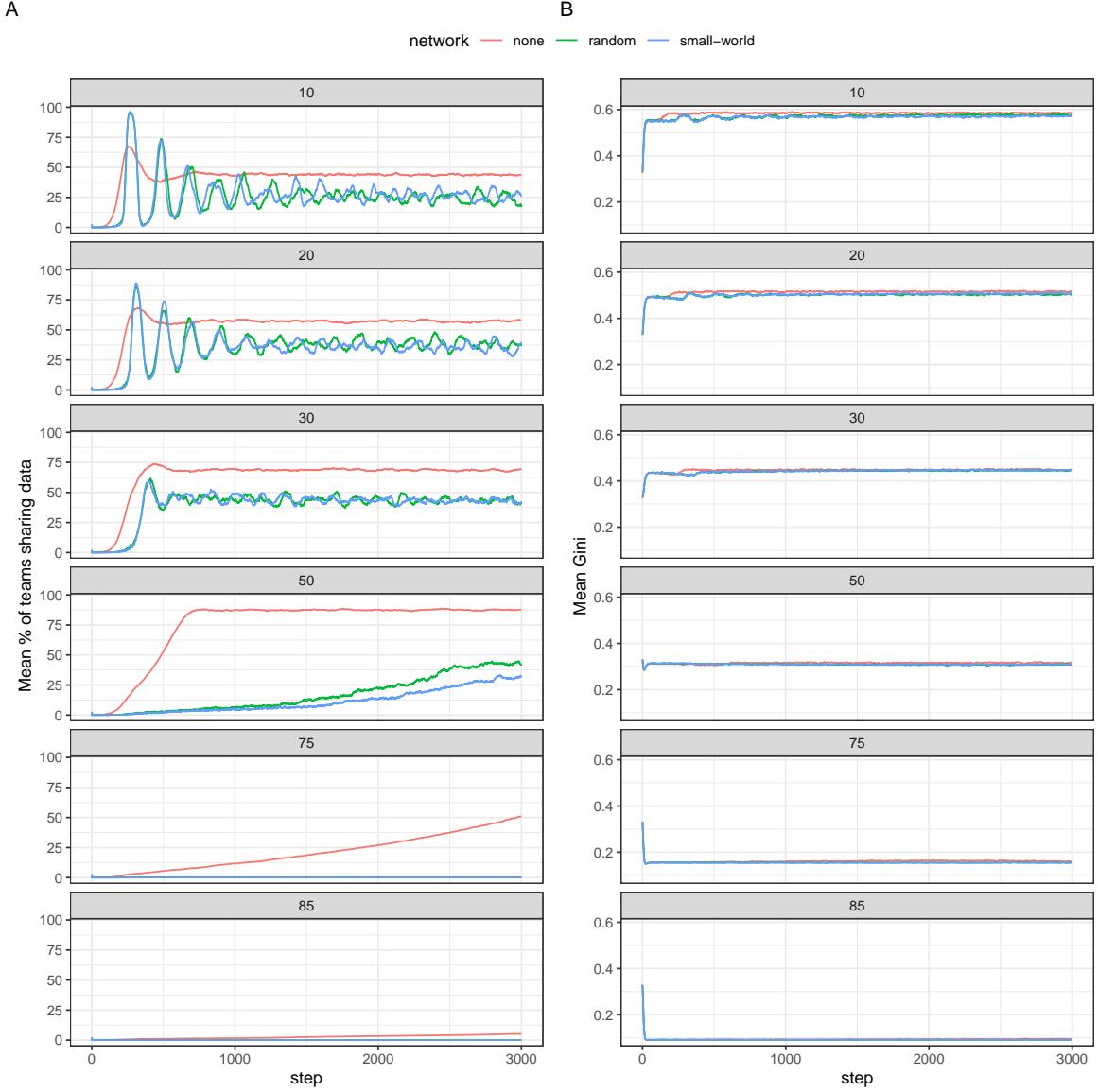


Figure 3: Effect of networks on (A) rate of sharing and (B) Gini coefficient. The rows represent the varying rate of funded teams in %. Uniform starting distribution.

The red lines in Figure 2 correspond to Figure 1. We can observe that the Gini is not substantially affected by different network structures, while the share of teams sharing data is strongly affected by the presence of network effects, but only weakly by the type of network.

Overall, network effects lead to a lower share of teams that share data, presumably because we start with no teams sharing data, and therefore the descriptive norms act as a dampener.

However, for larger grants (and a lower share of teams that is funded each round), the share of groups sharing data swings widely between two extreme points, before settling on a narrower equilibrium state.

Effect on success of different groups

```
group_success <- no_network %>%
  group_by(step, funded_share, init_dist) %>%
  summarise(across(contains("mean_funds"), .fns = mean))

pdata <- group_success %>%
  pivot_longer(contains("mean_funds"), names_to = "quantile",
               names_pattern = ".*_(q\\d)")

pdata %>%
  ggplot(aes(step, value, colour = quantile)) +
  geom_line() +
  facet_grid(rows = vars(funded_share),
             cols = vars(init_dist))
```

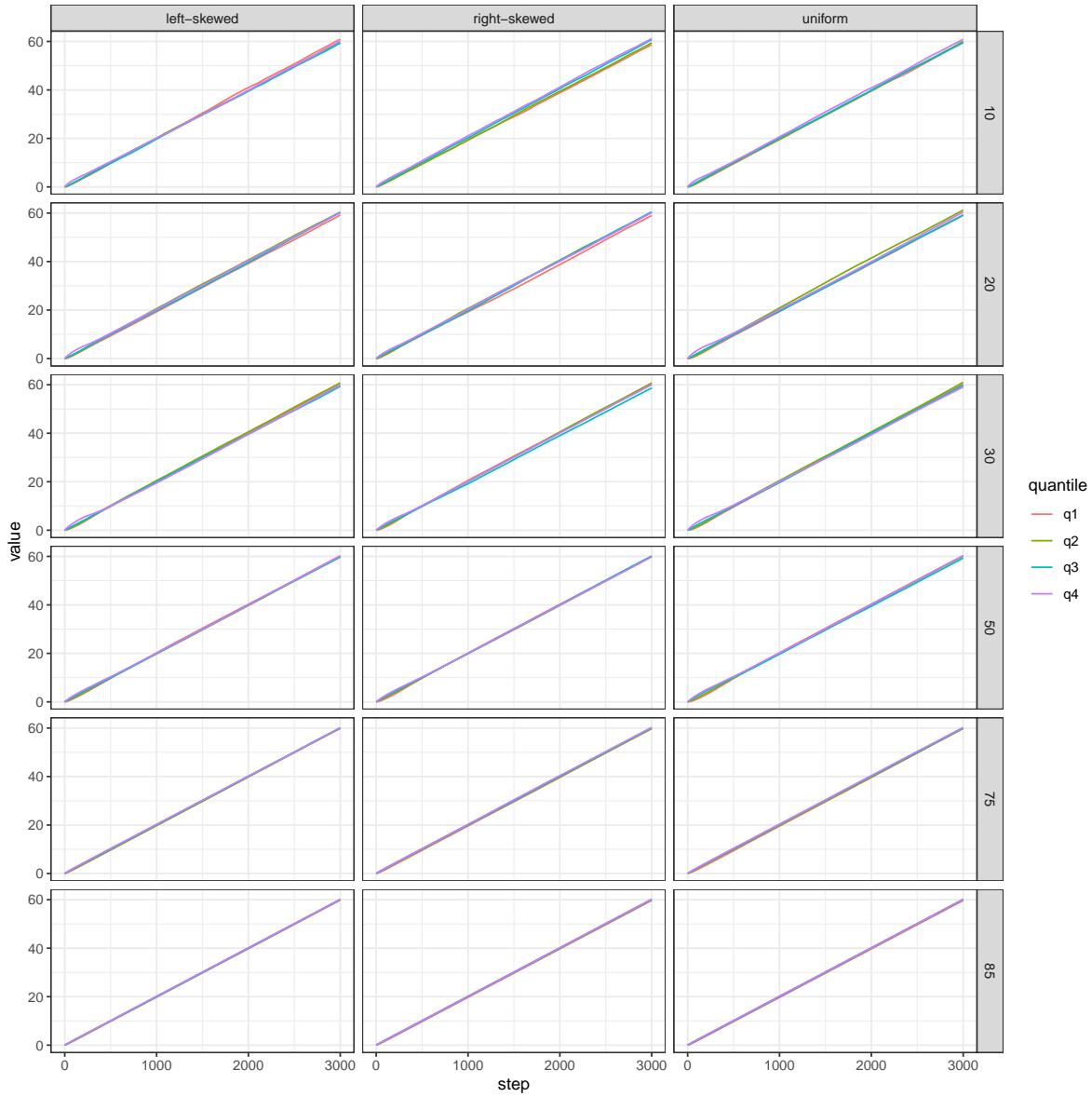


Figure 4: Mean resources by initial resource quantile with no network

There is no difference in how successful groups are based on their initial quantile, when there are not networks. Below we provide the same for a random network (Figure 5), and for a small-world network (Figure 6).

```
group_success <- df %>%
  filter(network == "random") %>%
  group_by(step, funded_share, init_dist) %>%
  summarise(across(contains("mean_funds"), .fns = mean))

pdata <- group_success %>%
  pivot_longer(contains("mean_funds"), names_to = "quantile",
               names_pattern = ".*_(q\\d)")

pdata %>%
  ggplot(aes(step, value, colour = quantile)) +
  geom_line() +
  facet_grid(rows = vars(funded_share),
             cols = vars(init_dist))
```

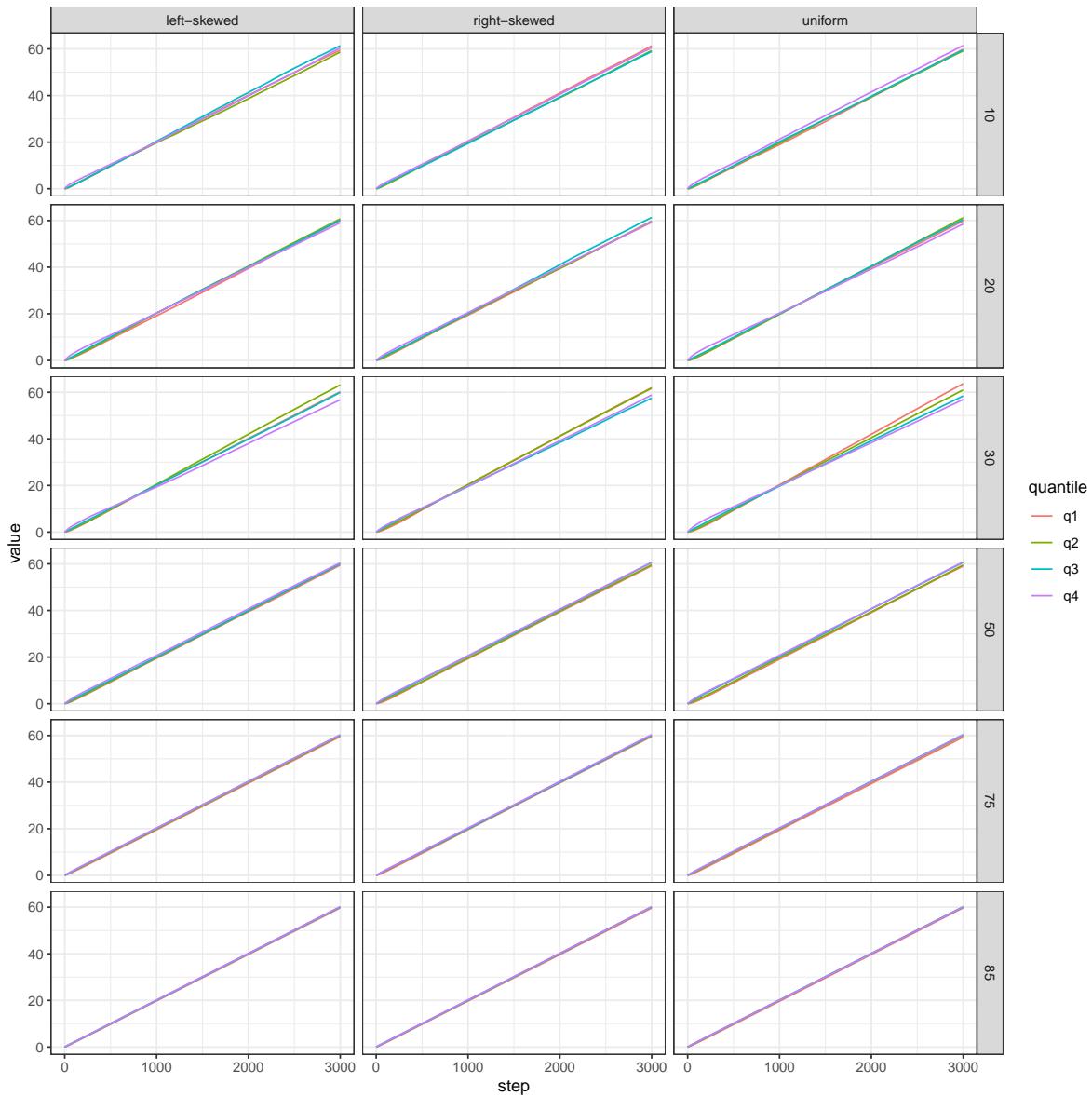


Figure 5: Mean resources by initial resource quantile with random network

```
group_success <- df %>%
  filter(network == "small-world") %>%
  group_by(step, funded_share, init_dist) %>%
  summarise(across(contains("mean_funds"), .fns = mean))
```

```
pdata <- group_success %>%
  pivot_longer(contains("mean_funds"), names_to = "quantile",
               names_pattern = ".*_(q\\d)")

pdata %>%
  ggplot(aes(step, value, colour = quantile)) +
  geom_line() +
  facet_grid(rows = vars(funded_share),
             cols = vars(init_dist))
```

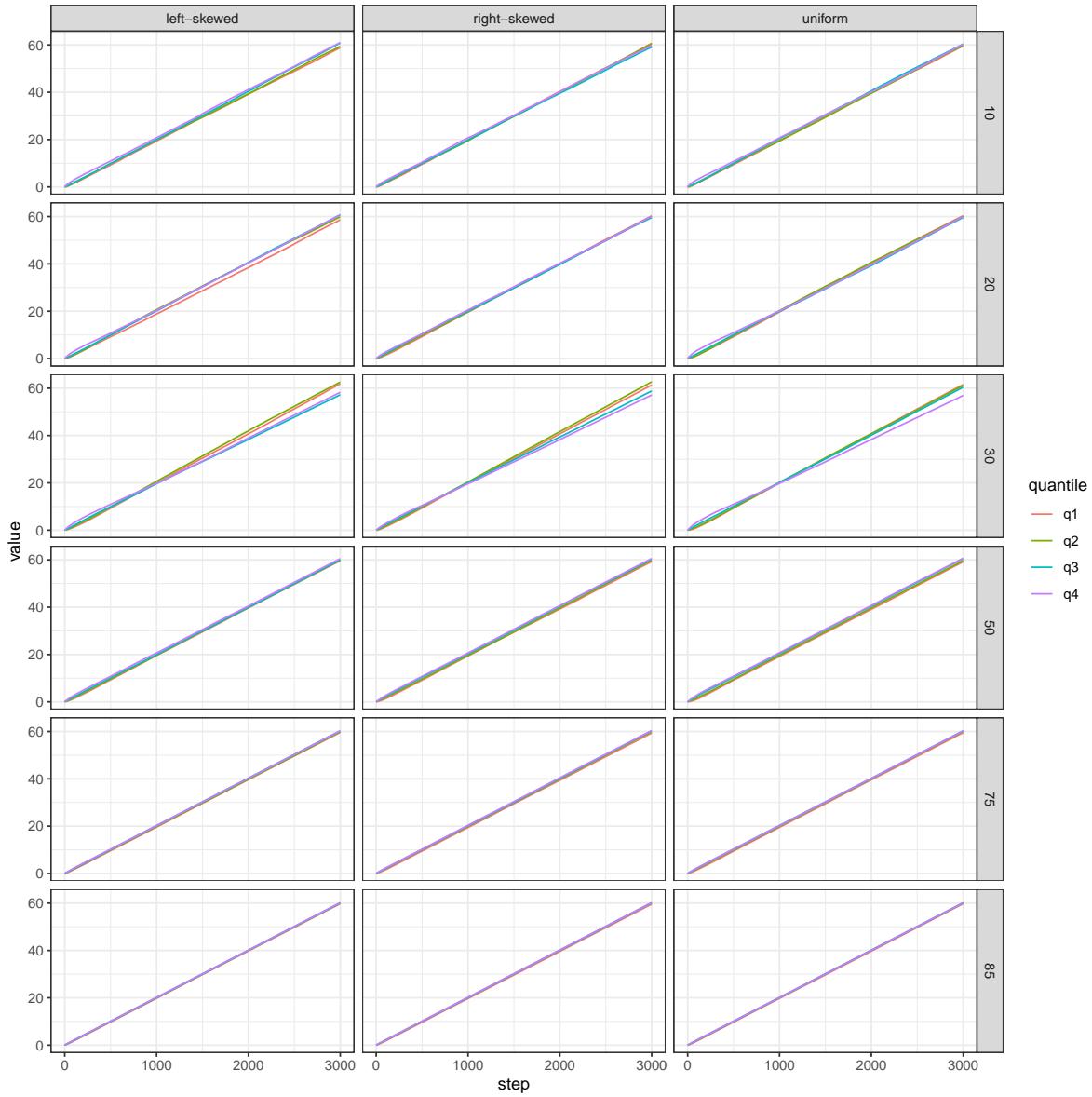


Figure 6: Mean resources by initial resource quantile with small-world network

In both cases there are slight differences, but I assume these are just random variation. Maybe the methodology of grouping teams based on their initial resources into four quartiles is not very useful - these are large groups that hide more fine-grained processes.